



REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD

Year 2005

Name of Author H. LEY, C. J.

COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOANS

Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

- A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).
- B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.
- C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.
- D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.



This copy has been deposited in the Library of

UCL



This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.

**AN INVESTIGATION INTO THE
RELATIONSHIP BETWEEN A PUTATIVE
TUMOUR SUPPRESSOR AND A CELL
SHAPE REGULATOR IN THE FRUIT FLY**

CHARLOTTE JANE HILEY

UNIVERSITY COLLEGE LONDON

PhD

UMI Number: U592878

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U592878

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I would like to thank my father, Robin Hiley, my mother, Carolyn Hiley, my sister, Emma Hiley, my granny, Margaret Whitcher, my boyfriend, Tim Landy, and all my friends for their help and support during my PhD. I would like to thank my supervisor, Kathy Barrett, for her guidance, and everyone past and present in Kathy's lab and in the Ludwig Institute for Cancer Research, UCL, who have helped me along the way.

In memory of my Mum.

Abstract

The process of morphogenesis during development involves complex, co-ordinated changes in cell shape and cell movement. The molecular mechanisms underlying this fundamental aspect of development remain to be fully elucidated, although some key aspects have been characterised.

Drosophila Rho Guanine nucleotide Exchange Factor 2 (DRhoGEF2) is required during morphogenesis for control of cell shape. A novel interaction has recently been discovered between the DRhoGEF2 PDZ domain and the *Drosophila* orthologue of Mutated in Colorectal Cancer (MCC), a putative tumour suppressor gene. The interaction between MCC and a RhoGEF has not been previously documented in any organism. MCC may act as a block on the cell cycle, and this interaction could, therefore, represent a link between morphogenesis and cell division. This thesis explores the functional significance of MCC and in particular its putative contribution to morphogenesis through its interaction with DRhoGEF2.

The observation of MCC expression in the *Drosophila* embryonic central nervous system suggests a functional role in nervous system development. Ectopic over-expression of MCC does not produce any obvious phenotype over controls. Furthermore, RNAi and P-element mutagenesis to knock out MCC expression do not yield a phenotype, and therefore do not indicate any clear function for MCC. The potential influence of MCC on the signals mediated by DRhoGEF2 therefore remains obscure and requires further investigation.

Three potential targets for the DRhoGEF2 PDZ domain identified through a yeast 2-hybrid approach, including MCC, carry a proline–X–threonine–X-leucine motif at their C-terminus. Initial studies indicate that mutation of the proline, threonine and leucine residues at the C-terminus of MCC

disrupts its binding to DRhoGEF2 PDZ domain. This work suggests further study to explore whether this C-terminal motif defines a group of DRhoGEF2-specific interacting proteins.

Table of Contents

List of Figures and Tables.....	11
Abbreviations	15
1. Introduction.....	19
1.1. The Rho GTPase family	19
1.2. The molecular switch.....	21
1.3. Rho GTPases have numerous effectors.....	22
1.4. Rho GEFs are major regulators of Rho GTPase activity	23
1.5. Rho GEFs may dictate the downstream outcome of Rho activation	25
1.6. Rho GTPases have numerous functions in tissue culture cells	25
1.6.1. Rho GTPases primarily mediate their effects via regulation of the actin cytoskeleton	26
1.6.2. Rho GTPases regulate cell migration	26
1.6.3. Rho GTPases regulate cell morphology	28
1.6.4. Rho GTPases are required for the establishment of cell polarity	30
1.6.5. Rho GTPases are required for cell proliferation.....	30
1.6.6. Rho GTPases are required for membrane trafficking	31
1.7. Rho and the fruit fly.....	32
1.7.1. Oogenesis	33
1.7.2. Pole cell formation	34
1.7.3. Cellularisation	34
1.7.4. Gastrulation	35
1.7.5. Epithelial regulation	36
1.7.6. Dorsal closure.....	38
1.7.7. Nervous system development	38
1.7.8. Muscle development.....	40
1.7.9. Eye development.....	40
1.7.10. Tissue polarity	40
1.7.11. Metamorphosis	41
1.8. <i>In vitro</i> studies complement <i>in vivo</i> studies.....	41
1.9. DRhoGEF2 functions in cellular contraction	41

1.10.	DRhoGEF2 is an RGS Rho GEF	42
1.11.	The RGS Rho GEF family participate in conserved signalling pathways	44
1.12.	RGS Rho GEFs are GAPs for heterotrimeric G proteins and GEFs for small G proteins	45
1.13.	RGS Rho GEFs have distinct expression patterns.....	48
1.14.	Human RGS Rho GEFs are recruited to the plasma membrane upon activation of G α protein signalling.....	49
1.15.	RGS Rho GEFs have effects on the actin cytoskeleton	50
1.16.	PDZRhoGEF and LARG interact with plexin B and mediate growth cone collapse.....	51
1.17.	LARG and possibly PDZRhoGEF are activated by the insulin-like growth factor 1 receptor	52
1.18.	LARG and PDZRhoGEF are regulated by phosphorylation ..	52
1.19.	RGS Rho GEFs can homo- and heterodimerise via their C-termini	53
1.20.	The Mutated in Colorectal Cancer (MCC) protein interacts with DRhoGEF2.....	54
1.21.	The MCC gene is mutated in a variety of human cancers.....	55
1.22.	Murine MCC is expressed in intestinal epithelia and nervous tissue	56
1.23.	MCC may play a role in cell cycle regulation.....	56
1.24.	MCC has a homologue, MCC-2	57
1.25.	The interaction between MCC and RGS Rho GEFs may provide a link between Rho GTPases and cancer.....	57
1.26.	Two other proteins also interact with the PDZ domain of DRhoGEF2.....	60
1.27.	PDZ domains are protein-protein interaction modules	61
1.28.	PDZ-containing proteins act as scaffolds and participate in signalling	61
1.29.	There are several mechanisms of interaction between PDZ domains and their targets	62

1.30.	The specificity of the PDZ-target interaction is determined by amino acid residues at the C-terminus of the target	63
1.31.	The PDZ domain of DRhoGEF2 is a class I PDZ domain	64
1.32.	Amino acids N-terminal to the -2 position are also important in the specificity of binding	64
1.33.	PDZ domains can be engineered to bind a specific C-terminal target sequence	66
1.34.	The hypothesis for this thesis	67
1.35.	The aims of this study	68
1.36.	Chapter 1 References	68
2.	Materials and methods	84
2.1.	Polymerase Chain Reaction (PCR)	84
2.2.	Restriction digestion and ligation of DNA	85
2.3.	Transformation of bacteria	85
2.4.	Diagnostic PCR	86
2.5.	Isolation of plasmid DNA from bacteria	86
2.6.	Generation of RNAi hybrid construct	86
2.7.	Microinjection of the hybrid RNAi construct to generate transgenic flies	86
2.8.	Selection and balancing of RNAi transgenic flies	87
2.9.	Transgenic RNAi and UAS-MCC crosses to GAL4 lines	87
2.10.	RNA extraction for RT-PCR	87
2.11.	Reverse transcription and PCR	88
2.12.	Probe generation for whole mount in situ hybridisation	89
2.13.	<i>In situ</i> hybridisation	89
2.14.	Plasmid rescue of P(lacW) insertions	90
2.15.	P-element hop genetics	91
2.16.	Genetic interaction studies	92
2.17.	Genomic DNA isolation for PCR screening	93
2.18.	Digoxigenin-DNA labelling of probes for Southern blotting ...	94
2.19.	Dot blot test	95
2.20.	DNA digestion and denaturation for southern blot	95
2.21.	Transfer of DNA to membrane and probe hybridisation	96

2.22.	Stringency washes and detection.....	96
2.23.	Storing, stripping and re-probing the membrane.....	97
2.24.	Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE).....	98
2.25.	Transfer and Immunoblotting	98
2.26.	Stripping and reprobing membrane.....	98
2.27.	Tagged DRhoGEF2 and MCC construct preparation.....	99
2.28.	Transfection and lysis of <i>Drosophila</i> cells	99
2.29.	Full-length protein co-immunoprecipitation	100
2.30.	Recombinant PDZ domain production	101
2.31.	PDZ domain co-immunoprecipitation	102
2.32.	Chapter 2 References	102
3.	Results – DMCC expression analysis	104
3.1.	Introduction.....	104
3.2.	The protein encoded by the predicted gene CG6156 interacts with DRhoGEF2 PDZ domain.....	104
3.3.	The predicted protein encoded by CG6156 has sequence homology to human MCC.....	106
3.4.	MCC has other species orthologues.....	107
3.5.	MCC has a conserved PDZ target motif at its C-terminus	107
3.6.	MCC proteins have a predicted repeat coiled coil structure....	111
3.7.	RT PCR analysis reveals that there are two additional <i>Drosophila</i> MCC isoforms.....	111
3.8.	DMCCc has five predicted coiled coils.....	115
3.9.	All four mRNAs are transcribed during embryogenesis	115
3.10.	MCC mRNA is expressed in the developing embryonic central nervous system	115
3.11.	Discussion and conclusions	119
3.11.1.	MCC gene and protein structure	119
3.11.2.	MCC mRNA expression	121
3.12.	Chapter 3 References	123
4.	Results – DMCC functional analysis.....	126
4.1.	Introduction.....	126

4.2.	The GAL4-UAS system	126
4.3.	Over-expression of MCC using the GAL4-UAS system.....	127
4.4.	Over-expression of MCC leads to no obvious developmental phenotype.....	127
4.5.	The use of RNA interference to compromise MCC function ...	128
4.6.	Generation of a UAS-MCC RNAi DNA construct.....	129
4.7.	Generation of transgenic UAS-MCC-RNAi flies and crossing to GAL4 drivers	131
4.8.	Flies expressing an MCC-RNAi construct are viable with no visible phenotype.....	132
4.9.	Strategy for generation of an MCC mutant	132
4.10.	P-elements are mobile transposons in <i>Drosophila</i> that can be used for mutagenesis	133
4.11.	The P-element starting point	133
4.12.	PCR screening of P-element insertions	135
4.13.	A single insertion into MCC.....	137
4.14.	P1 is not a “clean” insertion.....	137
4.15.	Molecular analysis of the P1 insertion.....	138
4.16.	The deletion does not extend as far as the easter gene	141
4.17.	Conclusions from the P1 line analysis.....	144
4.18.	Broadening of the PCR screening.....	144
4.19.	The P2 insertion may be incomplete	145
4.20.	The second hop	146
4.21.	Two distinct lines.....	147
4.22.	The P3/4 P-element may be incomplete	148
4.23.	Verifying the insertions	149
4.24.	The P1 line is viable over the P4 line	152
4.25.	The P1 and P4 lines complement deficiencies that take out MCC	153
4.26.	Fertility defects.....	154
4.27.	Discussion and conclusions	156
4.27.1.	RNAi and UAS-MCC expression studies	156
4.27.2.	P-element hopping (general comments)	158

4.27.3.	Southern blotting (general comments)	158
4.27.4.	P1 insertion analysis	159
4.27.5.	P4 insertion analysis leading to the conclusion that an MCC mutant has no obvious phenotype	162
4.27.6.	Complementation analysis	164
4.27.7.	Lack of phenotype.....	165
4.28.	Chapter 4 References	166
5.	Results – DMCC and DRhoGEF2 interaction analysis	168
5.1.	Introduction.....	168
5.2.	Interaction between the isolated DRhoGEF2 PDZ domain and DMCC168	
5.3.	Conservation of PDZ-binding motif	170
5.4.	Requirement of C-terminal residues for interaction	170
5.5.	Bioinformatics to identify further PXTXL proteins	171
5.6.	Interaction between full-length DRhoGEF2 and DMCC.....	172
5.7.	Genetic interaction between DRhoGEF2 and DMCC	173
5.8.	Discussion and conclusions.....	180
5.8.1.	Molecular interaction between DRhoGEF2 and DMCC...	180
5.8.2.	PDZ-target interaction specificity	185
5.8.3.	Modelling PDZ domain interactions	189
5.8.4.	Additional targets.....	192
5.8.5.	Genetic interactions.....	199
5.9.	Chapter 5 References.....	199
6.	Summary, conclusions and future work.....	202
6.1.	Summary of results.....	202
6.2.	MCC function.....	203
6.3.	DRhoGEF2 PDZ domain function.....	205
6.4.	Chapter 6 References.....	208
	Appendix 1: DNA and protein sequences.....	210
	Appendix 2: Primer sequences.....	221
	Appendix 3: TXL-COOH genes of <i>Drosophila</i> (PATSCAN result)	223

List of Figures and Tables

Figure 1.1. The human Rho GTPase family.....	20
Figure 1.2. Rho GTPase activators and effectors.	21
Figure 1.3. Rho GTPases are molecular switches.	22
Figure 1.4. Rho GTPases are required for cell migration.....	27
Figure 1.5. DRhoGEF2 is required for cell shape changes during gastrulation.	37
Figure 1.6. DRhoGEF2 putative signalling domains.	43
Figure 1.7. The RGS Rho GEF Family.	44
Figure 1.8. Putative orthologous signalling pathways.	46
Table 1.1. The RGS Rho GEFs.	47
Figure 1.9. The predicted signalling and protein interaction domains of PDZ Rho GEF.....	51
Figure 1.10. The PDZ-target interaction.....	63
Table 1.2. The classification of PDZ domains.	65
Table 2.1. GAL4 drivers.	88
Figure 3.1. CG6156 mRNAs.	105
Figure 3.2. Regions of sequence identity between human MCC1 and <i>Drosophila melanogaster</i> MCCa protein.....	106
Table 3.1. DMCC potential orthologues.	109
Figure 3.3. MCC similarity across species.	109
Figure 3.4. Fly and human MCC both have a classic C-terminal PDZ target sequence: T-X-L-COOH.	110
Figure 3.5. MCC has a repeat coiled coil structure.	110
Figure 3.6. Additional MCC mRNA isoforms.	112
Figure 3.7. DMCC mRNAs.....	113
Figure 3.8. Lack of expression of further MCC mRNAs.	114
Figure 3.10. MCC mRNA expression.....	116
Figure 3.11. BDGP embryonic microarray expression data for CG6156.	117
Figure 3.12. MCC may be expressed in the embryonic gut.	118
Figure 3.13. MCC expression in the central nervous system.....	119

Figure 3.14. CG6156 expression from embryo to adult.	120
Figure 3.15. The development of the central nervous system.	123
Figure 4.1. Over-expression of DMCC and MCC1 in <i>Drosophila</i>	128
Figure 4.2. Over-expression of MCC specifically in the <i>Drosophila</i> eye.	129
Table 4.1. Ubiquitous over-expression of MCC.....	129
Figure 4.3. Hybrid genomic DNA:cDNA construct.....	130
Figure 4.4. Generation of the genomic (A) and reverse cDNA (B) pieces of the RNAi hybrid construct.	131
Table 4.2. MCC RNAi expression studies.....	132
Figure 4.5. Location of the original insertion (P0).....	134
Figure 4.6. The P0 bristle phenotype.....	135
Figure 4.7. P(lacW).....	135
Figure 4.8. Primer positions for PCR screening.....	136
Table 4.3. 96-well PCR screening.....	137
Figure 4.9. P1 insertion discovery.....	138
Figure 4.10. P1 insertion position.....	138
Figure 4.11. Molecular mapping of the P1 P-element deletion.	139
Figure 4.12. Determining the extent of deletion of the P-element in P1 and investigating a possible 3' genomic deletion.	140
Figure 4.13. Molecular mapping of the P1 insertion line deletion.....	140
Figure 4.14. Strategy for sequencing to the right of the P1 insertion to determine the extent of deletion.....	142
Figure 4.15. Ligation and PCR to the right of the P-element in P1 line.	143
Table 4.4. easter complementation analysis.....	143
Figure 4.16. Broad screening genomic PCR primers.....	144
Figure 4.17. P2 insertion discovery.....	145
Figure 4.18. The P2 insertion.....	145
Figure 4.19. Discovery of the P3 insertions.	146
Figure 4.20. Isolation of the P3 and P4 insertion lines.	147
Figure 4.21. The P3 and P4 homozygous phenotype.	148
Figure 4.22. Testing for the presence of the original P-element insertions.	149

Figure 4.23. The P-element insertions as predicted by PCR analysis. The P0 insertion is not shown.	150
Figure 4.24. Probes for the Southern blot analysis.	150
Figure 4.25. Prediction of approximate wild-type band sizes for the Southern blot analysis.	151
Figure 4.26. Southern blot of P-element insertions using the MCC probe.	152
Figure 4.27. Southern blot of P-element insertions using the 3'gen probe.	153
Table 4.5. P-element complementation analysis.....	154
Figure 4.28. Complementation analysis of P-element lines.	155
Table 4.6. Fertility crosses for P-element insertions.	156
Figure 4.29. Polymorphism to the left of the MCC gene.	159
Figure 4.30. Hypothesis for P1 and P4 insertion outcomes and resulting Southern bands.	161
Figure 5.1. DMCCa co-immunoprecipitates with DRhoGEF2 PDZ domain.....	169
Table 5.1. C-terminal amino acid residues of DRhoGEF2 predicted PDZ targets.....	170
Figure 5.2. The proline and leucine residues are required for binding to DRhoGEF2 PDZ domain.	171
Table 5.2. PXTXL-COOH proteins of the Drosophila proteome (BLAST analysis).	172
Figure 5.3. Full-length DRhoGEF2 does not co-immunoprecipitate with DMCC in S2 cell lysates.	173
Table 5.3. DRhoGEF2 alleles.	174
Figure 5.4. DRhoGEF2 null / hypomorph phenotypes.	175
Table 5.4. Combination of heterozygous DRhoGEF2 with heterozygous DMCC.....	176
Table 5.5. Combination of homozygous DRhoGEF2 with heterozygous DMCC.....	177
Figure 5.5. Phenotypes of flies homozygous mutant for DRhoGEF2 and heterozygous mutant for DMCC.	179

Table 5.6. Combination of homozygous DRhoGEF2 with homozygous DMCC.....	182
Figure 5.6. Model for DRhoGEF2 interaction with DMCC.....	183
Figure 5.7. Model of the Shank PDZ domain interacting with a peptide target.	189
Figure 5.8. PDZ-target interactions.	191
Figure 5.9. PDZ domain primary sequence line-ups.	192
Figure 5.10. Caps protein structure.....	194
Figure 5.11. Embryonic gene expression profiles of DRhoGEF2 and the six PXTXL genes	198

Abbreviations

AIE – autoimmune enteropathy

AIEBP - autoimmune enteropathy binding protein (also known as MCC2)

Ala - alanine

APC – adenomatous polyposis coli

Arg – arginine

Asn - asparagine

Asp – aspartic acid

BDGP – Berkeley *Drosophila* genome project

BLAST – basic local alignment search tool

bp – base pair

BSA – bovine serum albumin

C.elegans – *Caenorhabditis elegans*

C-terminus – carboxyl-terminus

Caps - capricious

Cdc42 – Cell division cycle 42

CDK – cyclin-dependent kinase

CDM – Ced-5 / DOCK180 / Myoblast city

cDNA – complementary DNA

CNS – central nervous system

CSPD - Disodium 3-(4-meth-oxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo [3.3.1.1^{3,7}] decan}-4-yl)phenyl phosphate

Cta - Concertina

Cys - cysteine

CZH2 – CDM/zizimin homology 2

DAG - diacylglycerol

Dbl – diffuse B-cell lymphoma

ddH₂O – double-distilled water

DEPC - diethyl pyrocarbonate

DH – Dbl homology

Dia - Diaphanous

DIG – digoxigenin

DMCC – *Drosophila* mutated in colon cancer

D.melanogaster – *Drosophila melanogaster*

DNA – deoxyribonucleic acid

DRhoGEF – *Drosophila melanogaster* Rho guanine nucleotide exchange factor

Drosophila - *Drosophila melanogaster*

DTT - Dithiothreitol

EB1 – End-binding protein 1

EGFR – epidermal growth factor receptor

ERBIN – ErbB2 interacting protein

ERK – EGF receptor kinase

EMT – epithelial-mesenchymal transition

eye - eyeless

FAK – focal adhesion kinase

FAP – familial adenomatous polyposis

FGF – fibroblast growth factor

Fog – Folded gastrulation

FRET – fluorescence resonance energy transfer

FTI – farnesyl transferase inhibitor

GEF – guanine nucleotide exchange factor

GAP – GTPase activating protein

GDI – guanine nucleotide dissociation inhibitor

GDP – guanosine diphosphate

Gln – glutamine

Glu – glutamic acid

Gly – glycine

GMR – glass multimer reporter

GPCR – G-protein coupled receptor

GTPase – guanosine triphosphatase

GTP – guanosine triphosphate

His - histidine

HRP – horse radish peroxidase

HsMCC – homo sapiens mutated in colorectal cancer

Htl – Heartless

IGF – insulin-like growth factor
 IGFR – insulin growth factor receptor
 IP – immunoprecipitation
 IPTG - Isopropyl bD-thiogalactopyranoside
 Iso - isoleucine
 JNK – Jun N-terminal kinase
 Kb - kilobase
 LARG – leukemia associated Rho GEF
 LB – Luria Bertani
 Leu - leucine
 LPA – lypophosphatidic acid
 LRR – leucine-rich repeat
 Lys – lysine
 MAPK – mitogen-activated protein kinase
 Mbc – Myoblast city
 MCC – Mutated in colorectal cancer
 MDCK – Madine Derby Canine Kidney
 MEKK - mitogen-activated protein kinase/ERK kinase kinase
 Met – methionine
 MLCK – myosin light chain kinase
 MLL – mixed lineage leukaemia
 MTOC – microtubule organising centre
 N-terminus – amino-terminus
 NGF – nerve growth factor
 NHERF – Na/H⁺ exchange regulatory factor
 OD – optical density
 PAGE – polyacrylamide gel electrophoresis
 PAK – p21-activated kinase
 Pbl – Pebble
 PBS – phosphate buffered saline
 PCR – polymerase chain reaction
 PDZ - Postsynaptic density protein/Discs large/Zonula occludens
 PH – pleckstrin homology

Phe – phenylalanine
PI3K - phosphoinositide-3 kinase
PIP2 – phosphatidylinositol-3,4-diphosphate
PNS – peripheral nervous system
Pro - proline
Rac – Ras-related C3 botulinum toxin substrate
RGS – regulator of G-protein signalling
Rho – Ras homologue
RNA – ribonucleic acid
RNAi – RNA interference
ROK – Rho kinase
Rpm – revolutions per minute
RT PCR – reverse transcriptase polymerase chain reaction
S.cerevisiae – *Saccharomyces cerevisiae*
SDS – sodium dodecyl sulphate
sev - sevenless
SH2 – Src homology 2
SH3 – Src homology 3
siRNA – short interfering RNA
SMART – simple modular architecture research tool
Smc – segregation maintenance chromosomes
SNA1 - α -1 syntrophin
SRF – serum response factor
Thr - threonine
TIAM - T-cell lymphoma invasion and metastasis
Trp – tryptophan
Tyr – tyrosine
UAS - upstream activating sequence
Val - valine
WASP – Wiskott-Aldrich Syndrome Protein

1. Introduction

During the development of an organism, signals transmitted from one molecule to another regulate important processes such as cell division and migration, apoptosis (programmed cell death), and morphogenesis (the generation of shape and structure). The signalling pathways that directly regulate these and other cellular processes are themselves highly regulated, providing reproducible programs of development. It is becoming increasingly clear that signalling pathways that were once thought of as linear leading to a given outcome (such as entry into the cell cycle) are intimately linked to signalling pathways regulating other processes. For example, in early fruit fly development the protein Tribbles places a temporary block on signals promoting cell division to provide time for morphogenetic events to occur before the next round of cell division begins (Grosshans and Wieschaus 2000). In this way the signalling pathways regulating cell division and morphogenesis are regulated in a co-ordinated manner and may be thought of not as linear pathways, but as part of a complex network of signals within the cell.

The molecular components of cellular signalling pathways are primarily proteins. One group of proteins involved in cellular signalling is the Rho GTPases.

1.1. The Rho GTPase family

Rho GTPases are found in all eukaryotic cells from yeast to humans. They form a subgroup of the Ras superfamily of small GTPases, which number over 60 in mammalian cells. There are 20 Rho GTPase family members in humans, three of which are well-characterised in tissue culture systems; these are RhoA, Rac1 and Cdc42 (Etienne-Manneville and Hall, 2002). The phylogenetic tree of human Rho GTPase family members (**fig. 1.1**) illustrates that the naming of members does not necessarily indicate their degree of sequence similarity. For example, RhoG is more closely related to Rac1 than to RhoA. The Rho

GTPase family in *Drosophila* is much smaller than in mammals, with only seven identified thus far.

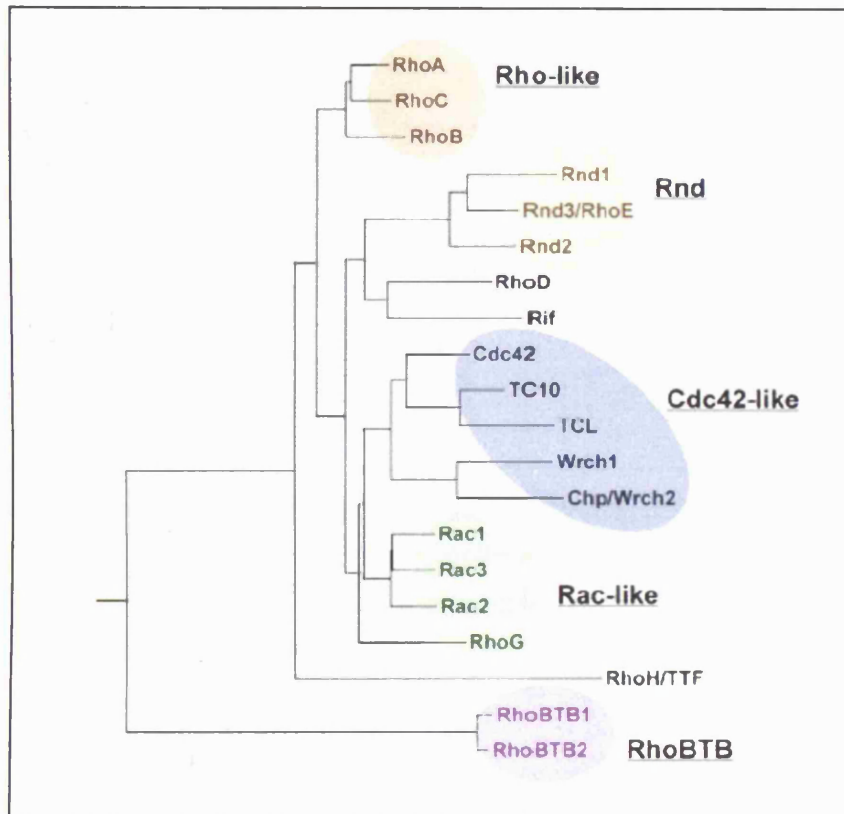


Figure 1.1. The human Rho GTPase family. – a phylogenetic tree. Taken from (Burridge and Wennerberg 2004).

Human Rho GTPases are liable to regulation by more than 70 activators and 80 inactivators. This large number of upstream regulators combined with the expanding number of Rho GTPase downstream targets that are continuously being identified (over 60 so far), indicates the complexity of Rho GTPase-mediated signalling (fig. 1.2).

Different Rho GTPases may have exclusive or over-lapping functions. However, the function of a particular Rho GTPase seems to be at least partially conserved through evolution since the best-characterised proteins (Rac, Rho and Cdc42) have similar functions in fruit flies and humans (see sections 1.6 – 1.7).

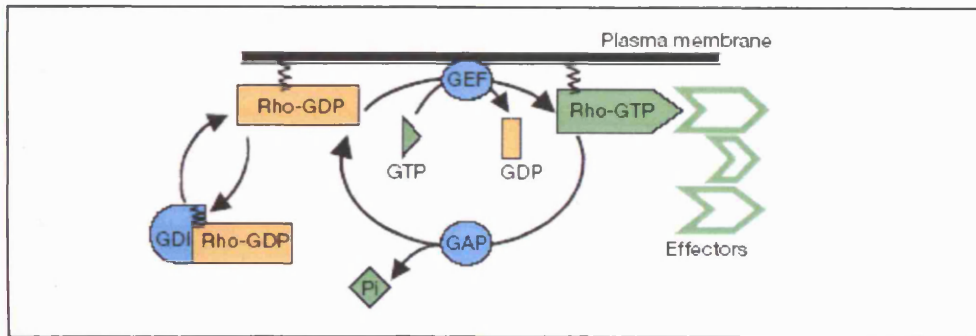


Figure 1.3. Rho GTPases are molecular switches. Rho GTPases are negatively regulated by GAPs and GDIs, and positively regulated by GEFs. Rho-GTP is the form that activates downstream effectors. Taken from (Etienne-Manneville and Hall 2002).

1.3. Rho GTPases have numerous effectors

Rho GTPases are involved in a diverse variety of cellular functions both *in vitro* and *in vivo*. This diversity of function is partially derived from the numerous downstream targets activated by Rho GTPases (fig. 1.2) (Bishop and Hall 2000). Since Rho GTPase effectors do not all contain a recognisable conserved motif they may have to be identified experimentally. A large number of identified effectors are kinases, including serine/threonine kinases (e.g. p21-activated kinase (PAK), mitogen-activated protein kinase (MAPK)/EGF receptor kinase kinase (ERKK)) and lipid kinases (e.g. phosphoinositide-3 kinase (PI3K)). The molecular mechanism by which Rho GTPases activate effectors is not entirely understood but is known to involve residues within a region termed the switch 1 domain as well as other regions of the Rho GTPase protein (Karnoub, Symons et al. 2004). Rho GTPase single amino acid mutations have begun to tease out differential effector functions. For example, the identification of Rac1 mutants that retain the ability to induce lamellipodia but have lost transforming activity implicate different effectors for these functions (Westwick, Lambert et al. 1997).

It is not clear why a Rho GTPase can activate one of its effectors and not another whilst it is itself in an active state. *In vivo* studies indicate that the Rho GEF activating the Rho GTPase may dictate the downstream effector, although the mechanism behind this is not understood (see section 1.5).

1.4. Rho GEFs are major regulators of Rho GTPase activity

Although Rho GTPase activation can be achieved by either GEF activation or inhibition of GAP or GDI activity, studies of oncogenic GEFs suggest that it is the exchange of guanine nucleotide that is the rate-limiting step. Therefore, localised control of GEF activity is extremely important in the regulation of Rho GTPases.

There are over 70 known Rho GEFs. Until recently, all Rho GEFs were found to contain Dbl (Diffuse B-cell lymphoma)-homology and pleckstrin homology domains which conferred their Rho GEF activity. The Dbl homology domain takes its name from a region of Dbl Rho GEF, the first Rho GEF to be identified. This region confers the ability to exchange GTP for GDP on Rho GTPases. The pleckstrin homology domain takes its name from a region of the pleckstrin protein that mediates localisation to the plasma membrane via interaction with phospholipids (Fuentes, Karnoub et al. 2003). Both domains are necessary for Rho GEF activity *in vivo*. However, members of the CDM (Ced-5, DOCK180, Myoblast city) family activate Rac, and zizimin activates Cdc42 despite a lack of Dbl homology domains (Braga 2002). In these cases the CZH2 (CDM zizimin homology 2) domain, which also mediates dimerisation, is required for activation of the GTPase (Meller, Irani-Tehrani et al. 2004).

The Dbl family GEFs, which form the vast majority of Rho GEFs, have three conserved helices that reorient the switch region of the GTPase to promote ejection of GDP. GTP can then take its place, and since GTP is more abundant in the cell than GDP this is the likely outcome, leading to activation of the Rho GTPase (Snyder, Worthylake et al. 2002).

It is only recently that the mechanism of specificity of a GEF for one Rho GTPase over another has begun to be elucidated. GEF promiscuity is variable, with some GEFs able to activate more than one Rho GTPase, e.g. Vav and Vav2 can activate RhoA/B/G, Rac1 and Cdc42, and others specifically activating only one e.g. Tiam-1 is Rac1-specific (Erickson and Cerione 2004). The basis for this specificity of activation is unravelling with the determination of crystal structures and mutational analyses. Whether a GEF activates Rac or Cdc42 is dependent on its ability to interact with the residue at position 56 of the GTPase – mutation of tryptophan in Rac1 to phenylalanine, the corresponding residue in Cdc42, prevents Tiam1 from activating Rac1, but enables a Cdc42-specific GEF, Intersectin-1, to activate it (Karnoub, Worthylake et al. 2001). The 56 position may not be important for RhoA activation since RhoA has a Trp56 residue, as Rac, but Tiam1 does not activate RhoA. The Asp45 and Glu54 residues may prove to dictate GEF specificity for RhoA based on initial studies (Snyder, Worthylake et al. 2002; Oleksy, Barton et al. 2004). Studies of this type have determined that the β 2- β 3 region of Rho GTPases, a poorly conserved region amongst them, determines which GEFs can activate a particular GTPase and which cannot.

The elucidation of structural specificity for one GTPase over another will permit the use of GTPase-specific tools in studies of GEF function. For example, Dbs Rho GEF is able to activate both RhoA and Cdc42. A mutant form of Dbs Rho GEF that can activate RhoA but not Cdc42 leads to a transformed phenotype when over-expressed. This result, along with further studies, indicates that it is the activation of RhoA, rather than Cdc42, that leads to the transformed phenotype (Cheng, Rossman et al. 2002). Similar studies may also shed light on some of the confusing phenotypes observed with the use of Rho GTPase dominant negative mutants. A dominant negative Rho GTPase may inhibit GEFs that are able to activate itself *and* other Rho GTPases, and therefore lead to non-GTPase specific effects.

1.5. Rho GEFs may dictate the downstream outcome of Rho activation

It is clear that the activation of a specific Rho GTPase can produce radically different outcomes (see sections 1.6.2-1.6.6). For example, RhoA activation can lead to cell migration, cell cycle regulation, or gene expression. Some evidence points towards Rho GEFs determining the outcome of Rho GTPase activation. Co-expression studies in mammalian cells reveal different outcomes of Rac or Cdc42 activation when different GEFs are over-expressed (Zhou, Wang et al. 1998). Recent work in *Drosophila* illustrates the reiteration of a specific outcome of Rho activation by a single GEF (Nikolaidou and Barrett 2004). In this case DRhoGEF2 activates Rho1 leading to epithelial sheet folding at multiple stages of development, but DRhoGEF2 is not thought to be required for other processes involving Rho1 such as cytokinesis, which requires a different Rho GEF, Pebble. Similarly, the Rac GEF Trio is required for axonal guidance and growth, but not for myoblast fusion which are all Rac-dependent processes (Hakeda-Suzuki, Ng et al. 2002). This differential outcome of activation of a given Rho GTPase could be the result of differential Rho effector expression patterns at different stages of development, differential subcellular localisation or potentially a molecular mechanism whereby activation by a specific GEF produces a differential molecular conformation of the Rho GTPase that dictates effector activation. At present there is no evidence for the latter theory.

1.6. Rho GTPases have numerous functions in tissue culture cells

The vast majority of information about the functions of Rho GTPases since they were first identified in 1985 comes from studies of their function in mammalian tissue culture cells.

1.6.1. Rho GTPases primarily mediate their effects via regulation of the actin cytoskeleton

Of the 24 Rho GTPase family members in mammalian cells, three are well-characterised. Rho A, Cdc42 and Rac1 all function primarily by regulating the actin cytoskeleton, although other functions have also been assigned. In short, Rho A is required for the assembly of actin-myosin filaments (stress fibres), Cdc42 is required for the formation of finger-like actin protrusions (filopodia), and Rac is required for the formation of actin-rich sheet-like protrusions (lamellipodia) (Ridley and Hall 1992; Ridley, Paterson et al. 1992; Kozma, Ahmed et al. 1995; Nobes and Hall 1995). The actin cytoskeleton provides the structural framework of the cell (along with the microtubules and intermediate filaments). In this respect it dictates the shape (i.e. physical structure) of a cell. Since the actin cytoskeleton is dynamic, it also plays a vital role in changes in cell morphology and cell migration.

1.6.2. Rho GTPases regulate cell migration

Rho GTPases act as regulators for both independent cell movement, and the co-ordinated cell movement of an epithelial layer. A single mammalian cell in tissue culture requires Rac to form actin-rich membrane protrusions at the leading edge and make new adhesions with the substratum (Small, Stradal et al. 2002), Cdc42 to control the direction of protrusion with regard to extracellular cues (Etienne-Manneville and Hall 2003), and Rho to stimulate actin-myosin contraction in the cell body and retraction of membrane at the rear of the migrating cell (Worthylake, Lemoine et al. 2001).

How does the cell actually process forward? The active GTPases are localised to specific areas of the migrating cell which allows the cell to regulate the actin cytoskeleton differentially at the front ("leading edge") and rear.

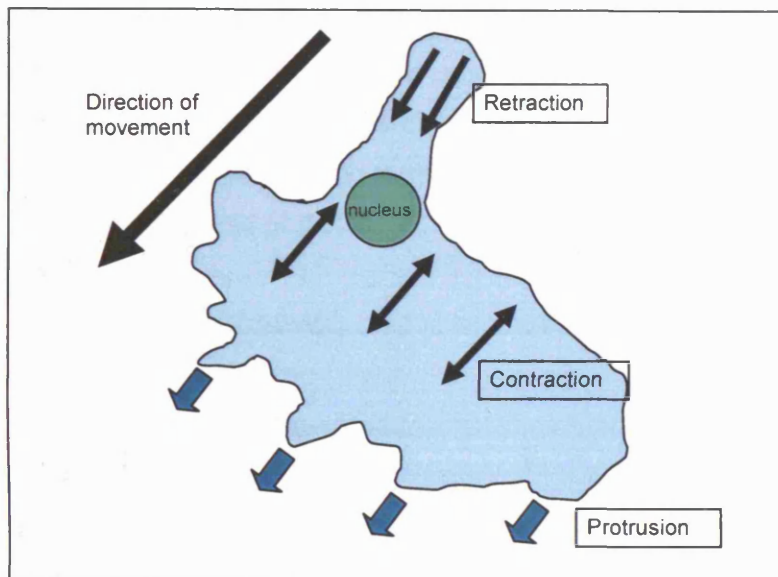


Figure 1.4. Rho GTPases are required for cell migration.

Rho proteins are required for protrusion at the leading edge, contraction in the main body of the cell, and retraction of the tail during migration. Based on (Etienne-Manneville and Hall 2002).

Rac function is required in the leading edge to promote the formation of lamellipodia which make new adhesions with the underlying substratum and allow the cell to move forward (Nobes and Hall 1995). Rho mediates its effects on tail retraction via phosphorylation of non-muscle myosin by myosin light chain kinase, which is regulated by Rho-dependent kinase (ROK). Inhibition of this pathway leads to inhibition of cell constriction and tail retraction, but protrusion at the leading edge is not affected (Kolega 2003). RhoA and ROK activity are also required to inhibit integrin-mediated adhesion to allow tail retraction (Worthylake, Lemoine et al. 2001). Although most studies to date have focussed on Rho A in mammalian cells, it is likely that Rho B also signals to ROK and thereby contributes to cell migration (Conway, James et al. 2004).

For a cell to process forward, protrusions must be co-ordinated to give rise to uni-directional migration (**fig. 1.4**). In a migrating cell Cdc42 dictates the direction of protrusion by interaction with atypical protein kinase C (aPKC). This interaction leads to the phosphorylation of

glycogen synthase kinase 3 (GSK-3) specifically at the leading edge of the cell. Gsk-3 phosphorylation promotes interaction of the adenomatous polyposis coli (APC) protein with the plus ends of microtubules (Etienne-Manneville and Hall 2003). In this way, the microtubule cytoskeleton becomes polarised and dictates the direction of protrusion.

Although the vast majority of studies have focussed on the role of Rho GTPases in regulation of the actin cytoskeleton, during cell migration the microtubule cytoskeleton is highly polarised, and new roles for Rho GTPases in the regulation of this part of the cytoskeleton are beginning to be unravelled. Rho is necessary for the accumulation of stabilised detyrosinated microtubules at the leading edge via integrin-mediated activation of focal adhesion kinase (FAK) (Palazzo, Eng et al. 2004). Rac inactivates the microtubule destabilising protein, stathmin, and thus encourages microtubule growth at the leading edge (Daub, Gevaert et al. 2001). Cdc42 is required for the orientation of the microtubule organising centre (MTOC) which is localised in front of the nucleus with respect to cell migration (Etienne-Manneville and Hall 2001).

Cells also move in a co-ordinated manner, for example as epithelial sheets. These are highly polarised structures which separate distinct extracellular spaces by virtue of junctions between neighbouring cells. Cells in a migrating epithelium must, therefore, move whilst attached to their neighbours, and Rho GTPases are also required for this process. In mammalian cell scratch assays inhibition of Cdc42 leads to random protrusive activity and cells do not migrate with normal efficiency (Etienne-Manneville and Hall 2001). It is also clear from *in vivo* studies of migrating epithelial sheets that other Rho GTPases are also required (see section 1.7.5).

1.6.3. Rho GTPases regulate cell morphology

Rho GTPases are required in the maintenance of integrity of epithelial and endothelial layers of cells. Cells within an epithelial layer exhibit a

cuboidal shape and are intimately connected to their neighbours via adherens junctions, which act as a mechanical link, and tight junctions, which act as a physical barrier. The development and maintenance of this specialised cell morphology is critical to the function of the epithelium. RhoA, Rac1 and Cdc42 are activated by, and required for, the formation of cadherin-cadherin cell contacts, which precede the formation of adherens junctions (Braga 2000). These cadherin-mediated cell contacts are initially brought about by filopodia and/or lamellipodia catalysing inter-cellular membrane contact (Vasioukhin, Bauer et al. 2000). Within an endothelium, a specialised form of epithelium that lines blood vessels, RhoA and Rac1 regulate permeability via effects on the formation of actin stress fibres and intercellular gaps (Wojciak-Stothard, Potempa et al. 2001).

Rho GTPases are also involved in the regulation of morphology of isolated cells. The changing shape of neuronal growth cones, which enables them to find their targets *in vivo*, provides a good system for study of Rho GTPase involvement in cell morphology. The axonal growth cone undergoes morphological changes in response to attractive and repulsive cues that direct it towards its target. A recent study showed that the plexin receptor, which is involved in repellent axon guidance via its interaction with semaphorin ligands, acts as a GAP for the small GTPase, Ras, which mediates the repulsive signal. However, this requires the interaction of Rnd1, a Rho GTPase, with the plexin receptor (Oinuma, Ishikawa et al. 2004). The RhoA GTPase is also required for semaphorin-mediated growth cone collapse via its regulation by PDZRhoGEF and Leukaemia Associated Rho GEF (LARG) which interact with plexin B1 (Aurandt, Vikis et al. 2002; Perrot, Vazquez-Prado et al. 2002; Swiercz, Kuner et al. 2002). Rac is required for growth cone collapse mediated by the semaphorins acting through plexin, and in this case (plexin A1) it appears that Rac is acting upstream of the plexin receptor (Turner, Nicholls et al. 2004). These are recent examples of the

many studies that highlight the importance of Rho GTPases in the process of axonal guidance.

1.6.4. Rho GTPases are required for the establishment of cell polarity

Cell polarity is important in unicellular organisms during chemotaxis and budding, and for various cell types in multicellular organisms such as neurons that receive a signal at one end of the cell and send it onwards at the other.

Without Cdc42 function, the yeast *Saccharomyces cerevisiae* cannot establish a defined site for daughter cell growth during the budding phase of the cell cycle, and cells therefore grow isotropically (Pruyne and Bretscher 2000). In multicellular organisms, Cdc42 plays a role in the development of apical/basolateral polarity within an epithelium via its interaction at the tight junction with Par proteins and atypical protein kinase C (Lin, Edwards et al. 2000). This polarity-determining role extends to migrating epithelial cells where Cdc42, once again in conjunction with aPKC and Par-6, regulates the positioning of the microtubule organising centre (MTOC) and the direction of protrusive activity (Etienne-Manneville and Hall 2001). Thus Cdc42 is required for the determination of cell polarity from yeast to mammalian cells.

1.6.5. Rho GTPases are required for cell proliferation

Rho proteins are required for cell cycle progression in tissue culture cells. Rho, Rac and Cdc42 all promote entry into G1, and progression into S-phase when expressed in quiescent fibroblasts. Inhibition of any of the three using either dominant negatives or toxins blocks G1 progression (Olson, Ashworth et al. 1995). The way in which the Rho GTPases exert their effects on cell cycle progression appears to be cell-type specific and is not fully understood (Coleman, Marshall et al. 2004). Rho GTPases are known to affect cyclin-dependent kinase (CDK) inhibitor levels and cyclin D levels. RhoA inhibits expression of the CDK inhibitors

p21^{Waf1/Cip1}, p27^{kip1} and p16^{Ink4} to promote G1-S phase transition. Cyclin D levels are regulated by two separate pathways. Upon mitogenic stimulation Ras activation promotes an increase in cyclin D levels, and Rho GTPases receive inputs from integrins that determine adhesion status in order to increase cyclin D levels at the correct time. Rac and Cdc42 are likely to be involved in a separate signalling pathway affecting cyclin D levels that is Ras-independent.

RhoE is a member of the Rho GTPase family that receives little attention, perhaps because it is not able to hydrolyse GTP and is therefore unusual among Rho GTPases. However, it does inhibit progression of the cell cycle at S-phase entry by preventing accumulation of cyclin D1, and this effect is independent of RhoA (Villalonga, Guasch et al. 2004).

Cdc42 regulates the attachment of microtubules to chromosomal kinetochores during mitosis (Yasuda, Oceguera-Yanez et al. 2004). This is a further example of Rho GTPases regulating the microtubule cytoskeleton. RhoA and Cdc42 are also both required for cytokinesis, the separation of daughter cells. Inhibition or constitutive activation of RhoA or Cdc42 inhibits the formation of a contractile actin-myosin ring required for cytokinesis (Glotzer 2001).

Since Rho GTPases regulate the cell cycle it is perhaps not surprising that they are implicated in the progression of cells to a cancerous state (see section 1.25).

1.6.6. Rho GTPases are required for membrane trafficking

Membrane trafficking involves the transport of membrane, proteins, and vesicular contents from one part of the cell to another. In polarised cells proteins may be trafficked specifically to only one plasma membrane compartment. The process of endocytosis involves the formation of vesicles at the plasma membrane which then move along the endocytic pathway to their final destination, commonly the lysosome. RhoB and

RhoD are implicated in the regulation of endosomal trafficking (Qualmann and Mellor 2003). RhoD regulates the interactions of early endosomes with the actin cytoskeleton. RhoB (and not RhoA) is involved in the trafficking of the epidermal growth factor receptor (EGFR) along the endocytic pathway to the lysosome.

Phagocytosis may be considered a specialised form of endocytosis whereby foreign material is ingested via a membrane-bound vesicle. Two phagocytic pathways in mammalian macrophages require the involvement of distinct Rho GTPases. Type I phagocytosis, which can occur through activation of the immunoglobulin receptor, requires Rac and Cdc42, whereas type II activation, which can occur through the complement receptor, requires Rho (Caron and Hall 1998). The involvement of Rho GTPases in phagocytosis by professional phagocytes of the immune system helps explain why they are such common targets for bacterial toxins.

Cdc42 has also been shown to regulate protein transport from the Golgi apparatus to the endoplasmic reticulum (Luna, Matas et al. 2002). Exactly how Cdc42 is mediating these effects is unclear, but it does appear to require Wiskott-Aldrich syndrome protein (WASP) and the actin cytoskeleton.

1.7. Rho and the fruit fly

Tissue culture studies of Rho GTPase function in mammalian cells have elucidated many basic cellular functions and biochemical interactors for Rho GTPases. More recent work using model organisms such as the mouse, fruit fly, *Drosophila melanogaster*, and nematode worm, *Caenorhabditis elegans*, is now shedding light on the biological functions of Rho GTPases *in vivo*. Rho GTPases play important roles at numerous stages of organism development. Here, the known functions of Rho GTPases in *Drosophila* development will be discussed.

Drosophila is an enormously powerful tool for the study of cell biology as it is amenable to genetic manipulation. *Drosophila* has been studied in this way since the early 1900s and a great deal of information and reagents are available for use today. The complete sequencing of the *Drosophila* genome further strengthened the use of *Drosophila* in cell biology (Adams, Celniker et al. 2000). For these reasons, and the fact that genetics can be used to study interactions between proteins, *Drosophila* has been used in numerous studies of the function of Rho GTPases.

Seven Rho GTPases have been identified thus far in *Drosophila*: Rho1, Rac1, Rac2, RhoL (Rac3), Rho BTB, Mtl, and Cdc42 (Rivero, Dislich et al. 2001; Settleman 2001). All are expressed during embryogenesis, some with restricted expression (Rac1 and Cdc42), others ubiquitously (Mtl, Rho1 and Rac2). As with Rho GTPases in mammalian cells, most studies have focussed on the roles of Rho1, Rac1 and Cdc42, the closest orthologues of RhoA, Rac1 and Cdc42 respectively. Functions for Rho GTPases have been identified at various stages of development from oogenesis through to eye development. Many of these functions tie in with observations from tissue culture studies.

1.7.1. Oogenesis

During oogenesis the developing oocyte receives cytoplasm from the nurse cells to which it is connected via actin-rich ring canals. These germ cells are surrounded by a single layer of somatic follicle cells which are also connected to the nurse cells via actin-rich adherens junctions. Studies of constitutively active and dominant negative forms of RhoL, Rac1 and Cdc42 have implicated these proteins in the regulation of actin-rich connections between these different cell types (Murphy and Montell 1996). Studies of Rho1 maternal mutants illustrate that it is also required for proper ring canal function (Magie, Meyer et al. 1999) and Cdc42 loss of function mutants indicate that Cdc42 is required throughout

development to maintain monolayered epithelia, for example in the follicular cell epithelium during oogenesis (Genova, Jong et al. 2000).

As well as their function in regulating actin-rich cell connections, Rho GTPases may also be important in cell migratory events during oogenesis. A group of follicle cells termed border cells migrate towards the oocyte at the anterior tip of the egg chamber. A receptor-tyrosine kinase mediated signal is received by these cells to guide them on their journey. Rac1 and one of its GEFs, Myoblast city (Mbc), act downstream of the receptor and are likely to influence the actin cytoskeleton during the process of guided cell migration (Duchek and Rorth 2001).

1.7.2. Pole cell formation

The first cells to form in the *Drosophila* embryo are the pole cells, which constitute the germline. DRhoGEF2, an upstream activator of Rho1, is required for normal actin and myosin localisation within the pole cells, and in DRhoGEF2 mutants the majority of pole cells fail to pinch off from the somatic syncytium and are destroyed (Padash Barmchi, Rogers et al. 2005). In the formation of the pole cells, DRhoGEF2 appears to be functioning in the regulation of actin-mediated cellular contraction, a process in which it is repeatedly implicated during embryonic development (see below).

1.7.3. Cellularisation

Cellularisation is a specialised form of cytokinesis whereby membrane moves inwards between adjacent nuclei in the early embryo forming furrow canals. These furrow canals are actin and myosin-rich at their leading edge. Dominant negative Rho and constitutively active Cdc42 disrupt the formation of furrow canals between nuclei during cellularisation and phenocopy the effect of cytochalasin (an inhibitor of actin polymerisation) implying that it is their role in regulation of the actin cytoskeleton that is required (Crawford, Harden et al. 1998). Studies of Rho1, its upstream activator, Pebble (Pbl), and its downstream effector,

Citron kinase, implicate this signalling pathway in cytokinesis in every tissue examined (Shandala, Gregory et al. 2004). Thus Rho GTPases are required for cellularisation and cytokinesis throughout *Drosophila* development, corroborating their requirement during cytokinesis of tissue culture cells.

DRhoGEF2 and Diaphanous (Dia), a downstream effector of Rho1, may function in the same pathway during furrow canal formation since both are localised to forming furrows and are required for proper furrow formation (Grosshans, Wenzl et al. 2005). As for pole cell formation, this process involves actomyosin-mediated constriction. Dia is a member of the formin family that organises actin polymerisation (Waller and Alberts 2003). Dia localisation may be dependent on DRhoGEF2, and it is likely that Rho1 provides a link between the two proteins since both can interact with it (Grosshans, Wenzl et al. 2005). Both DRhoGEF2 and Dia have also been reported to interact with EB1, a microtubule associated protein (Rogers, Wiedemann et al. 2004; Wen, Eng et al. 2004). EB1 could, therefore, provide a second link between DRhoGEF2 and Dia.

1.7.4. Gastrulation

Gastrulation is a process common to multicellular organisms whereby cells become specified as endodermal, mesodermal or ectodermal. This process is accompanied by morphogenetic cell movements which result in the physical separation of these cell types. During *Drosophila* gastrulation cells along the ventral midline of the embryo invaginate to form the ventral furrow. Dominant negative studies suggest that these mesodermal precursor cells constrict their apical membranes in a Rho1-dependent manner. DRhoGEF2 orchestrates these cell shape changes (fig. 1.5) (Barrett, Leptin et al. 1997; Hacker and Perrimon 1998), probably via regulation of myosin II localisation (Nikolaidou and Barrett 2004).

DRhoGEF2 is thought to be functioning in a signalling pathway from the cell surface to the actin cytoskeleton. Embryos lacking Folded

gastrulation (Fog) (a putative extracellular ligand) or Concertina (Cta), a G α constituent of a heterotrimeric G protein, also fail to execute the cell shape changes necessary for ventral furrow formation (Parks and Wieschaus 1991; Costa, Wilson et al. 1994). Thus, a signalling pathway is postulated whereby Fog binds to an unknown G-protein coupled receptor (GPCR) and activates Cta, which in turn binds to and activates DRhoGEF2 leading to Rho1 activation, myosin relocalisation, actin cytoskeleton rearrangements and cell shape changes. This pathway seems to be conserved from flies to humans (see section 1.11).

Following gastrulation, the newly designated mesodermal cells undergo a transition from epithelium to mesenchyme in order to migrate internally along the ectoderm. The Fibroblast Growth Factor (FGF) receptor, Heartless (Htl), is required for these cell shape changes and for attachment of mesodermal cells to the ectoderm. Recent work has also indicated a role for the Rho GEF, Pbl, in the regulation of these cell shape changes (Schumacher, Gryzik et al. 2004). Unusually, Pbl appears to be acting independently of Rho1 since the expression of dominant negative Rho1 does not affect mesodermal cell shape changes.

1.7.5. Epithelial regulation

As for tissue culture cells, Rho GTPases are necessary in *Drosophila* development for the formation and maintenance of epithelial layers. As mentioned above, Cdc42 is required for the maintenance of a monolayered epithelium throughout development (Genova, Jong et al. 2000) and Rac is required for the formation of adherens junctions (Eaton, Auvinen et al. 1995). Rho acts antagonistically to moesin, an actin-binding protein, in the maintenance of integrity of an epithelium. Moesin appears to function by keeping Rho signalling “in check” thus preventing epithelial cells from losing their apical actin and migrating away (Speck, Hughes et al. 2003). It remains to be seen whether the same is true in mammalian cell epithelia.

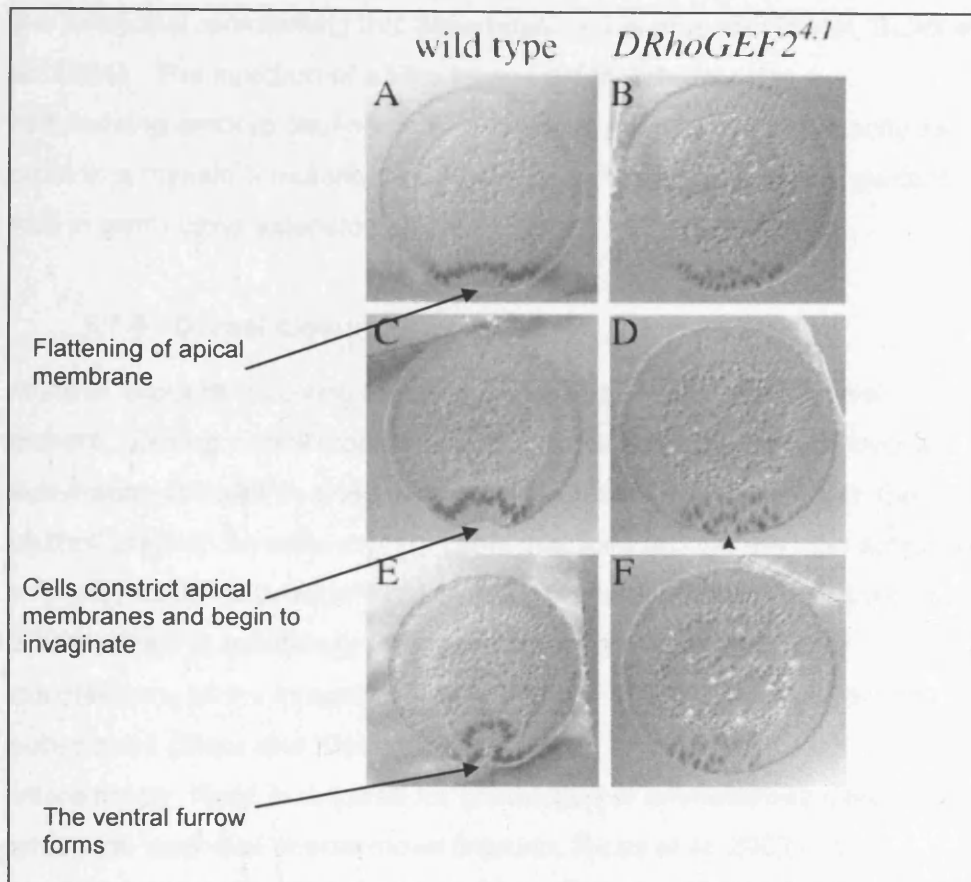


Figure 1.5. DRhoGEF2 is required for cell shape changes during gastrulation. Cross-sections of early embryos show that DRhoGEF2 mutants (*DRhoGEF2*^{4.1}) do not undergo the cell shape changes required for invagination of the ventral furrow. Occasionally a cell does constrict its apical membrane in the *DRhoGEF2* mutant (arrowhead) but there is no co-ordinated constriction. **A, C, E:** wild-type, **B, D, F:** *DRhoGEF2* mutant. Taken from (Barrett, Leptin et al. 1997).

Epithelia are not static structures; they must maintain integrity during the numerous morphogenetic movements they undergo during development. Rho GTPases play important roles in this maintenance. DRhoGEF2 is required during ventral furrow formation, but it is also required reiteratively during embryogenesis when an epithelial sheet is required to fold during a morphogenetic process, such as salivary gland formation (Nikolaidou and Barrett 2004). During germ band elongation, when epithelial cells intercalate to elongate the embryo, myosin II is required for

the junctional remodelling that accompanies this process (Bertet, Sulak et al. 2004). The injection of a Rho kinase (ROK) inhibitor into a cellularising embryo causes defects in junctional remodelling exactly as seen in a myosin II mutant, thus Rho is likely to be playing an important role in germ band extension.

1.7.6. Dorsal closure

Another process involving the morphogenesis of epithelia is dorsal closure. During dorsal closure sheets of epithelial cells migrate over a substratum in order to close a hole in the epithelium. Cells within the epithelium form an actin-myosin cable that runs around the hole acting as a “purse-string” to provide force to close the hole (Hutson, Tokutake et al. 2003). Rho1 is necessary for the formation of this cable and for maintenance of the integrity of the epithelium whilst it moves over the substratum (Bloor and Kiehart 2002; Jacinto, Wood et al. 2002). Interestingly, Rac1 is required for contraction of amnioserosa cells, over which the epithelial sheets move (Harden, Ricos et al. 2002). The leading edge cells within the migrating epithelium also send out lamellipodia and filopodia which are regulated by Cdc42 and Rac1, and enable cells to adhere to their correct partners in the final stages of dorsal closure (Jacinto, Wood et al. 2000).

The initiation of dorsal closure is regulated by the JNK (Jun N-terminal kinase) signalling cascade. Studies in mammalian systems show this cascade to be downstream of Rho GTPase signalling and there is evidence to support Rac and Cdc42 involvement in initiation of *Drosophila* dorsal closure, although it is unlikely that Rho1 is involved (Settleman 2001).

1.7.7. Nervous system development

As described above, the requirement for Rho GTPases in axonal guidance is well-established in tissue culture systems. Studies in

Drosophila confirm the requirement for Rho GTPases in axon guidance *in vivo*, but some results conflict with tissue culture studies. Contrary to studies with mammalian tissue culture neurons, in a *Drosophila* embryo it appears that Rho does not play a role in axon guidance since axons project normally in a fly mutant for Rho1 (Lee, Winter et al. 2000). However, dendrites are over-extended in this mutant, and the dendritic tree is reduced when Rho1 is over-expressed indicating that Rho1 may play a role in dendritic morphogenesis in fruit flies. The Rac GTPases are, however, required in axon guidance in *Drosophila*. Loss of all three Racs (Rac1, Rac2, Mtl) leads to defects in axon growth, guidance and branching, and it is likely that these effects are mediated via different Rac effectors (Ng, Nardine et al. 2002). The Rac GEF Trio mediates axonal guidance, and still life, another Rac GEF, is required for differentiation of axons into mature synapses (Sone, Hoshino et al. 1997; Hakeda-Suzuki, Ng et al. 2002).

At the *Drosophila* midline axons must make a decision whether to cross in order to reach their targets for innervation. A system of attractive and repulsive cues directs axons across the midline and prevents them re-crossing. Recent work indicates that a RhoGAP, Vilse, is required for the repulsive response mediated by the Robo receptor in axons crossing the midline (Lundstrom, Gallio et al. 2004). This RhoGAP promotes the intrinsic GTPase activities of Rac and, to a lesser extent, Cdc42, thereby inactivating them. These results, combined with genetic interaction data, suggest a Robo signalling pathway involving Rac inactivation mediates the repulsive response at the midline.

The nervous system is not made up solely of neurons – glia are required to ensheath axons in order for them to transmit electrical signals effectively. Glia in the *Drosophila* peripheral nervous system must first migrate from the lateral edge of the central nervous system to the periphery and then ensheath their target axons. A study of the actin cytoskeleton within these cells revealed dynamic rearrangements during

this migration process and the requirement for the small GTPases Rho1 and Rac1 for both migration and ensheathment of axons (Sepp and Auld 2003).

1.7.8. Muscle development

Muscles are formed in flies by the fusion of myoblasts into mature syncytial muscle fibres, as in mammals. The Rac triple mutant fly exhibits defects in myoblast fusion, confirming earlier studies with dominant negative Rac (Hakeda-Suzuki, Ng et al. 2002). In this case it is the Rac GEF Myoblast city (Mbc) that is probably responsible for the regulation of Rac activity.

1.7.9. Eye development

Drosophila have compound eyes made from over 800 ommatidia each composed of eight photoreceptor cells and twelve accessory cells. The over-expression of Rho1, Rac1, Rac2 or Cdc42 in the eye leads to a rough eye caused by the disruption of photoreceptor morphology or organisation (Hariharan, Hu et al. 1995; Nolan, Barrett et al. 1998). In the case of Rho1, this is a direct result of disorganised actin in the photoreceptors (Hariharan, Hu et al. 1995). Further studies with Rac indicate its involvement in the regulation of photoreceptors by rhodopsin. Activated Rac is able to rescue a rhodopsin mutant in which photoreceptors degenerate (Chang and Ready 2000).

1.7.10. Tissue polarity

Epidermal structures such as the eye are polarised along an axis orthogonal to their apico-basal axis. Rho acts downstream of the Frizzled G-protein coupled receptor (Strutt, Weber et al. 1997) to set up this planar cell polarity by rearranging the actin cytoskeleton via ROK-mediated myosin II activation (Winter, Wang et al. 2001). This pathway has also proved to be important in establishment of polarity and

convergent cell extension in the zebrafish (Marlow, Topczewski et al. 2002).

Studies with dominant negative Rac had indicated a function for this small GTPase in the establishment of tissue polarity (Fanto, Weber et al. 2000), but more recent work with loss of function triple Rac mutants contradicts this (Hakeda-Suzuki, Ng et al. 2002).

1.7.11. Metamorphosis

The transition from larva to pupa towards the end of adult development is regulated by the steroid hormone ecdysone. A downstream effector of Rho1, LIM kinase, regulates ecdysone-induced gene expression during this transition (Chen, Gajowniczek et al. 2004). In *Drosophila* tissue culture cells this response is dependent on the actin cytoskeleton and the SRF transcription factor. This is the first evidence of potential involvement of Rho GTPases in metamorphosis.

1.8. *In vitro* studies complement *in vivo* studies

Much of the *in vivo* evidence of Rho GTPase function during development complements results from mammalian tissue culture studies. However, *in vivo* mutant studies highlight the limitations of using dominant negative forms of Rho GTPases for study either *in vitro* or *in vivo* since dominant negatives sometimes produce non-specific phenotypes not seen in studies with mutants and must therefore be interpreted with caution.

1.9. DRhoGEF2 functions in cellular contraction

As stated above, DRhoGEF2 functions as a GEF for *Drosophila* Rho1 GTPase. The study of DRhoGEF2 function provides the backdrop to this thesis. DRhoGEF2 is required for cell shape changes during ventral furrow formation in *Drosophila* gastrulation, as mentioned above. Cells at the ventral midline constrict their apical membranes to allow invagination of the ventral furrow (fig. 1.5). DRhoGEF2 also induces contractile cell shape changes when over-expressed in *Drosophila* S2 cells in culture

(Rogers, Wiedemann et al. 2004). In both cases there is evidence that DRhoGEF2 is regulating myosin regulatory light chain localisation (Nikolaidou and Barrett 2004; Rogers, Wiedemann et al. 2004). It is likely that DRhoGEF2 activates DROK via Rho1. In mammalian cells ROK phosphorylates regulatory myosin light chain (Amano, Ito et al. 1996) activating acto-myosin contraction, and it is likely that DROK does the same.

It is also interesting to note that DRhoGEF2 associates with the plus ends of microtubules in S2 cells via an interaction with End-binding protein 1 (EB1) (Rogers, Wiedemann et al. 2004). Upon over-expression of a constitutively active form of Cta (the G α protein that is predicted to activate DRhoGEF2) DRhoGEF2 relocates from the microtubule tips probably to the plasma membrane and cells round up. It is currently unknown whether DRhoGEF2 is localised to microtubule tips *in vivo*, and if so how this affects DRhoGEF2 and myosin localisation in apically-constricting cells during ventral furrow formation.

1.10. DRhoGEF2 is an RGS Rho GEF

Analysis of the primary sequence of DRhoGEF2 indicates adjacent DH and PH domains that are presumed to confer Rho GEF activity. The *Drosophila* genome encodes 22 predicted Rho GEFs with DH and PH domains. Similarly to the mammalian system, it appears that some *Drosophila* Rho GEFs are promiscuous in their activation of Rho GTPases whereas others are more specific. Genetic interactions imply that DRhoGEF2 is able to activate Rho1, but not Cdc42 or Rac1 (Barrett, Leptin et al. 1997), and biochemical analysis indicates that DRhoGEF2 activates Rho1 but not Rac1, Rac2, RhoL (Rac3), Mtl, or Cdc42 (Grosshans, Wenzl et al. 2005).

The primary sequence of DRhoGEF2 also indicates the presence of other putative signalling domains (fig. 1.6). The regulator of G-protein signalling (RGS) domain is where the G α protein, Concertina, is believed

to bind and thereby activate GEF activity. The existence of this domain places DRhoGEF2 in the family of RGS Rho GEFs.

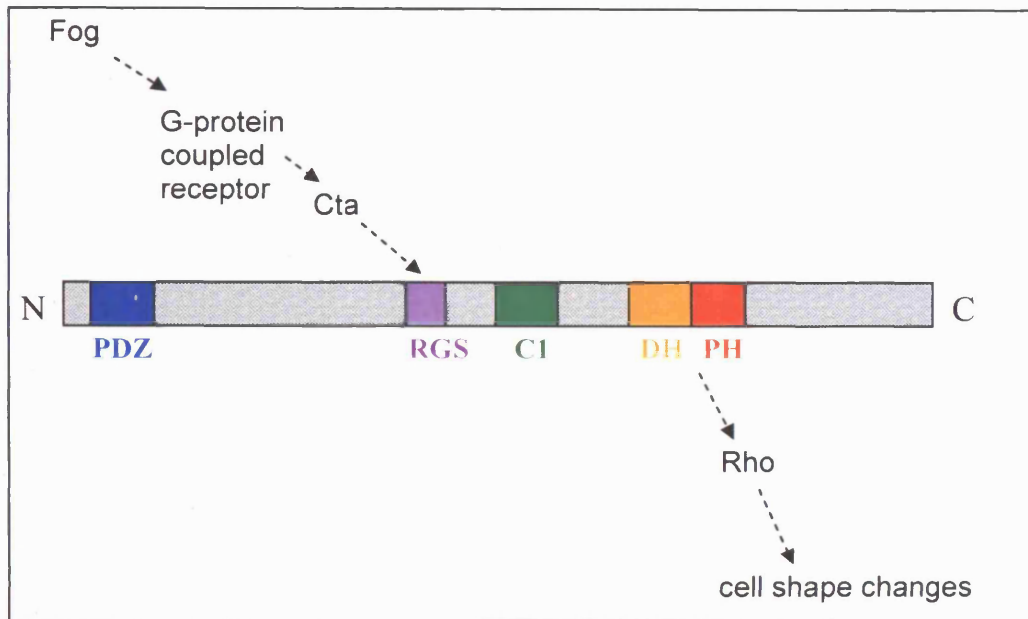


Figure 1.6. DRhoGEF2 putative signalling domains. DRho GEF2 is a large protein (~280KDa) with many putative signalling domains. PDZ: Postsynaptic density protein/Discs large/Zonula occludens domain, RGS: Regulator of G-protein signalling domain, C1: phorbol ester binding domain, DH: Dbl homology domain, PH: pleckstrin homology domain. N: N-terminus, C: C-terminus. Dashed lines indicate putative signalling pathway. Not to scale.

These proteins act as a link between heterotrimeric G-proteins and Rho GTPases. The human orthologues of DRhoGEF2, p115 Rho GEF, PDZRhoGEF and LARG, exchange GDP for GTP on RhoA GTPase in direct response to $G\alpha$ binding (Hart, Jiang et al. 1998; Rumenapp, Blomquist et al. 1999; Suzuki, Nakamura et al. 2003).

The primary sequence of DRhoGEF2 also indicates the presence of a putative phorbol ester / diacylglycerol (DAG)-binding or C1 domain. Protein kinase C contains such a domain and is directly activated upon binding diacylglycerol. The significance of this domain in DRhoGEF2 is not understood. The final putative signalling domain identified by primary

sequence analysis of DRhoGEF2 is a Postsynaptic density protein/Discs large/Zonula occludens (PDZ) domain. This will be discussed in detail below.

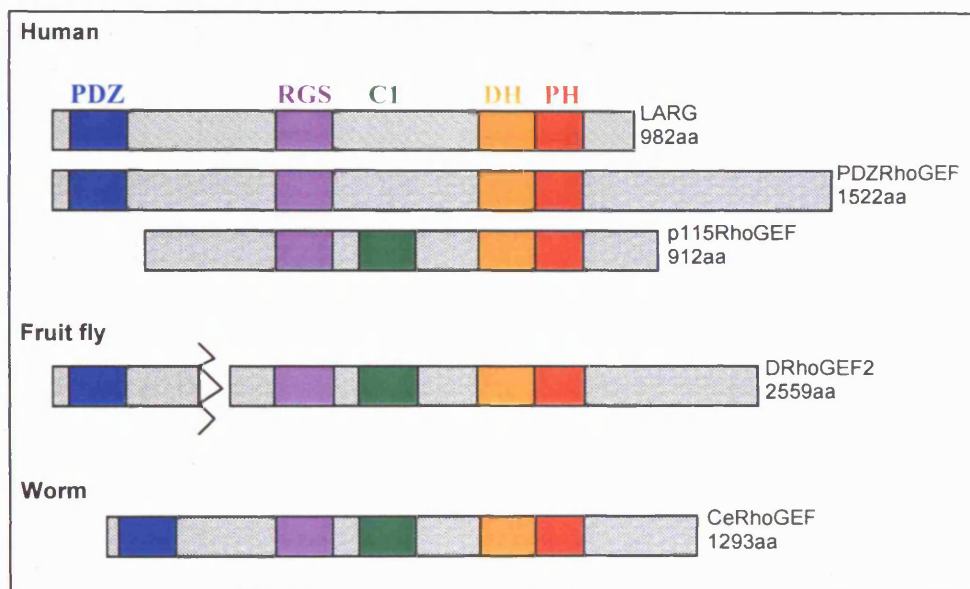


Figure 1.7. The RGS Rho GEF Family. LARG: leukaemia associated Rho GEF, PDZ: PDZ domain, RGS: RGS domain, PEB: phorbol ester binding domain, DH: Dbl homology domain, PH: pleckstrin homology domain. Numbers indicate protein size in amino acids. Not to scale.

1.11. The RGS Rho GEF family participate in conserved signalling pathways

The identification of DRhoGEF2 in *Drosophila* marked the beginning of the discovery of RGS Rho GEFs in various species. In humans the documented discovery of p115RhoGEF was closely followed by PDZRhoGEF (Fukuhara, Murga et al. 1999) which has a rat orthologue, GTRAP48, and LARG (Kourlas, Strout et al. 2000), which has also been cloned in mouse and *Xenopus* (Zinovyeve, Sveshnikova et al. 2004). Recently another member, CeRho GEF in *C.elegans*, has been added to the RGS Rho GEF family (fig. 1.7) (Yau, Yokoyama et al. 2003). The human orthologues of DRhoGEF2 each contain only a subset of the putative signalling domains, whereas CeRho GEF shares all predicted

domains with DRhoGEF2 indicating that these domains may have been lost in gene duplication events during evolution.

In all cases there is genetic or biochemical evidence that the RGS Rho GEFs participate in a signalling pathway involving their activation by a G α protein and their subsequent activation of a Rho GTPase. In the case of the human and *Drosophila* proteins there is also considerable genetic and biochemical data to support the existence of a signalling pathway from the cell surface via a GPCR to the actin cytoskeleton, resulting in the regulation of cell shape. In particular, in neurons it is well-established that lysophosphatidic acid (LPA) binding to its receptor initiates a signalling pathway via G $\alpha_{12/13}$ to induce actin cytoskeleton rearrangements, growth cone collapse and neurite retraction (Kranenburg, Poland et al. 1999). Now, with the functional assignment of PDZRhoGEF in neurite retraction it is possible to see how the whole pathway fits together, and a homologous pathway is likely to exist in the fruit fly (fig. 1.8) (Togashi, Nagata et al. 2000).

1.12. RGS Rho GEFs are GAPs for heterotrimeric G proteins and GEFs for small G proteins

p115RhoGEF exhibits GAP activity towards G α_{12} and G α_{13} *in vitro* (Kozasa, Jiang et al. 1998). The RGS domain increases the GTPase activity of the G α protein thereby generating a negative feedback on the signalling pathway. P115RhoGEF is also able to act as a GEF for Rho but not Rac, Cdc42 or Ras small GTPases (Hart, Sharma et al. 1996). The combined GAP/GEF activity is likely to reflect precise regulation of the signalling pathway from GPCRs through to Rho GTPases via a negative feedback system. LARG and PDZRhoGEF have been shown to act as Rho GEFs for RhoA, but not Rac1 or Cdc42 (Rumenapp, Blomquist et al. 1999; Taya, Inagaki et al. 2001).

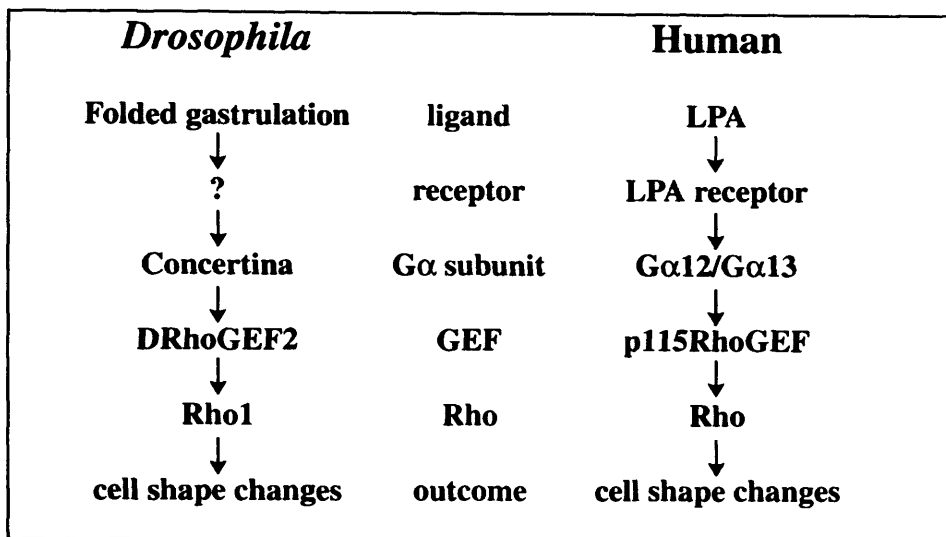


Figure 1.8. Putative orthologous signalling pathways. Signalling pathways from an extracellular ligand to the actin cytoskeleton are likely to be orthologous in *Drosophila* and humans. Adapted from (Barrett, Leptin et al. 1997).

LARG GAP activity has been demonstrated for both Gα₁₂ and Gα₁₃, and, similarly to p115RhoGEF, its GAP activity is restricted to Gα_{12/13} of all the heterotrimeric G proteins tested (Suzuki, Nakamura et al. 2003).

The specificity of Gα proteins in the activation of RGS Rho GEFs is currently under debate. Cta is a Gα protein related to both Gα₁₂ and Gα₁₃ - the *Drosophila* genome is predicted to encode only one Gα protein of this class. There is no evidence to suggest that DRhoGEF2 is activated by other Gα proteins, although this has not been directly tested to date. Many of the human RGS Rho GEFs can interact with more than one G protein in co-immunoprecipitation experiments, but this does not necessarily imply that each of those G proteins activates that particular RGS Rho GEF. Previous studies indicate that solely Gα₁₂ is able to activate LARG, but more recent studies also implicate Gα_q (Booden, Siderovski et al. 2002; Chikumi, Fukuhara et al. 2002). This may be dependent on the cell type, and it should be noted that the direct stimulation of LARG by Gα_q *in vitro* remains to be shown. p115RhoGEF can be activated by Gα₁₃, but Gα₁₂ inhibits this activation (Hart, Sharma et

al. 1996). The RGS domain of GTRAP48 (the rat orthologue of PDZRhoGEF) is able to activate RhoA when it is itself activated by constitutively active $G\alpha_{13}$, but it only weakly interacts with $G\alpha_{12}$ (Wells, Liu et al. 2002). These results are summarised in Table 1.1.

	Ligand specificity	$G\alpha$ specificity	Interactors	Functions
LARG	LPA?	12 q??	IGF-1 receptor Plexin B FAK Tec LARG PDZRhoGEF	Axon growth cone regulation Actin cytoskeleton regulation
PDZRhoGEF/ GTRAP48	Thrombin?	13 12??	Actin EAAT4 PAK4 Plexin B FAK PDZRhoGEF LARG	Neurite retraction Axon growth cone regulation Actin cytoskeleton regulation
p115RhoGEF/ Lsc	LPA? sphingosine 1- phosphate?	13	p115RhoGEF	B and T cell immune response Actin cytoskeleton regulation
DRhoGEF2	Folded gastrulation?	Concertina?	MCC Mec2 group 3 EB1	Actin cytoskeleton regulation Cell contraction

Table 1.1. The RGS Rho GEFs. Upstream activators, interactors and downstream effects for the RGS Rho GEFs.

Different extracellular ligands may ultimately activate different RGS Rho GEFs. The use of short interfering RNAs (siRNAs) has illustrated that lypophosphatidic acid (LPA) stimulation of Rho signalling requires LARG whereas thrombin stimulation of Rho signalling requires PDZRhoGEF in prostate cancer cells (Wang, Liu et al. 2004). In both cases the addition of siRNA to p115RhoGEF has no effect, although p115RhoGEF isolated from fibroblast cells stimulated with LPA or sphingosine 1-phosphate is more active than when isolated from unstimulated cells (Wells, Gutowski et al. 2001).

It has recently been proposed that thrombin and LPA receptors activate different G proteins, $G\alpha_{12}$ and $G\alpha_{13}$ respectively (Yamaguchi, Katoh et al. 2003). Given this, and the fact that LARG is activated by $G\alpha_{12}$ and PDZRhoGEF by $G\alpha_{13}$, it is tempting to speculate that there may be two distinct pathways using different receptors, G α proteins and RGS Rho GEFs, that both ultimately converge on Rho. However, caution should be exercised since these relationships have been documented in only one cell type to date.

1.13. RGS Rho GEFs have distinct expression patterns

DRhoGEF2 protein is expressed primarily in epithelia during embryogenesis in *Drosophila* and is localised to the apical end of cells. Expression levels are elevated in tissues undergoing morphogenetic processes involving folding of membranes, and in the central nervous system (Padash Barmchi, Rogers et al. 2005). An extensive analysis of the expression of PDZRhoGEF, p115RhoGEF and LARG proteins has been carried out in murine tissues. p115RhoGEF (or Lsc as the murine orthologue is known) is primarily expressed in haematopoietic cells in the mouse (Girkontaite, Missy et al. 2001). LARG is expressed across a wide variety of tissues including brain, smooth muscle, spleen, intestinal epithelium and skin (Becknell, Shen et al. 2003). PDZRhoGEF, by contrast, is expressed at high levels only in the brain (Kuner, Swiercz et al. 2002).

The expression of PDZRhoGEF and LARG in the murine nervous system has been further analysed in conjunction with expression analysis of $G\alpha_{12/13}$ (Kuner, Swiercz et al. 2002). Although both PDZRhoGEF and LARG are widely detected in the brain and spinal cord, a striking difference in subcellular localisation is observed. LARG is enriched in cell bodies of neurons whereas PDZRhoGEF localises to neuronal processes. This fits with the observed interaction between PDZRhoGEF and the neuronal glutamate transporter, EAAT4, which is expressed at synaptic junctions of glutaminergic neurons (Jackson, Song et al. 2001). Interestingly, $G\alpha_{12}$ primarily co-localises with LARG in cell bodies whereas $G\alpha_{13}$ co-localises with PDZRhoGEF in neuronal processes. This could represent the proposed specificity of RGS activation by $G\alpha$ proteins, as described above. All four proteins are also expressed in dorsal root ganglia of the peripheral nervous system.

1.14. Human RGS Rho GEFs are recruited to the plasma membrane upon activation of $G\alpha$ protein signalling

p115RhoGEF localises to the cytoplasm of resting cells and relocates to the plasma membrane upon activation by LPA (Wells, Gutowski et al. 2001). In epithelial cells LARG is primarily localised to lateral membranes with a small proportion cytoplasmically localised. PDZRhoGEF is localised to the plasma membrane or peri-plasma membrane in resting fibroblast cells (Togashi, Nagata et al. 2000). These and other studies are difficult to compare since they make use of different cell types and some study endogenous proteins, others over-expressed proteins. However, a consensus seems to be that p115RhoGEF is primarily cytoplasmically localised, whereas PDZRhoGEF and LARG are primarily plasma membrane or peri-plasma membrane localised in a resting cell. Upon stimulation of GPCR signalling (e.g. by LPA addition) the RGS Rho GEFs localise to the plasma membrane.

1.15. RGS Rho GEFs have effects on the actin cytoskeleton

In mammalian fibroblasts over-expression of LARG results in the formation of actin stress fibres. This is abrogated by the co-expression of a dominant negative RhoA (Taya, Inagaki et al. 2001). p115RhoGEF also induces the formation of actin stress fibres when over-expressed (Togashi, Nagata et al. 2000). B and T cells derived from an Lsc knockout mouse show a profound reduction in actin polymerisation and improper migration compared to cells from a wild-type mouse (Girkontaite, Missy et al. 2001).

The over-expression of PDZRhoGEF in fibroblasts induces a relocalisation of actin to the cortical area beneath the plasma membrane and the cells round up. PDZRhoGEF interacts with actin via a short stretch of amino acids between the RGS and DH domains (Banerjee and Wedegaertner 2004). Two groups have studied the requirement for the different domains of PDZRhoGEF in the induction of actin-mediated cell shape changes in various cell lines. According to Togashi et al, the DH and PH domains are necessary and sufficient to induce stress fibre assembly, but the presence of a proline-rich domain C-terminal to the PH domain (fig. 1.9) is required to induce cell rounding and cortical actin assembly. Banerjee and Wedegaertner show that the actin-binding domain mediates localisation of PDZRhoGEF to the cell cortex but it may serve to inhibit Rho signalling since an actin-binding mutant shows increased levels of cell rounding. In neither case was a requirement for the PDZ domain in cellular localisation observed.

Studies of endogenous PDZRhoGEF localisation in a neuronal cell line indicate that it relocalises to the tips of retracting neurites upon addition of lypophosphatidic acid (LPA) (Togashi, Nagata et al. 2000). This mirrors the relocalisation of Rho upon LPA stimulation indicating that PDZRhoGEF could be interacting with Rho at neurite tips to promote actin reorganisation. Further evidence is provided by over-expression of

a dominant negative PDZRhoGEF that lacks the DH/PH domains and inhibits LPA-induced neurite retraction and cell rounding.

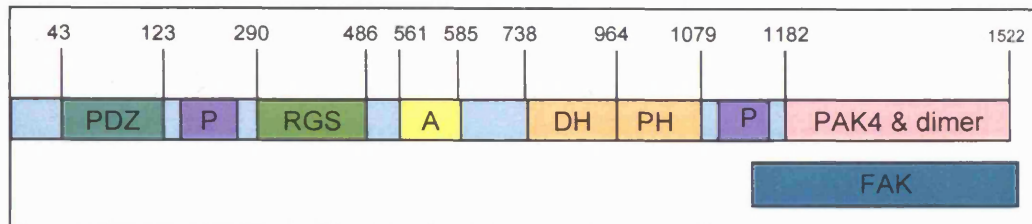


Figure 1.9. The predicted signalling and protein interaction domains of PDZ Rho GEF. PDZ: PDZ domain, P: proline-rich region, RGS: RGS domain, A: actin-binding domain, DH: DH domain, PH: PH domain, PAK4: PAK4 binding domain, dimer: dimerisation domain, FAK: domain of FAK phosphorylation which overlaps with PAK4 binding and dimer domains. Numbers refer to published amino acid sequence. Not to scale.

1.16. PDZRhoGEF and LARG interact with plexin B and mediate growth cone collapse

The direct interaction between plexin B and RGS Rho GEFs provides the missing link between plexin activation and Rho-mediated growth cone collapse (Swiercz, Kuner et al. 2002). Plexins are expressed on the surface of axon growth cones and mediate their targeting via interaction with semaphorin ligands that guide the axon. Plexin B1/2/3 share a canonical PDZ binding motif at their C-terminus (Thr-X-Leu-COOH) and the interaction is mediated via the PDZ domain of LARG / PDZRhoGEF. The interaction enables the plexin receptor to come to the membrane and respond to the semaphorin ligand since PDZ Rho GEF and LARG are localised at the membrane. The interaction leads to a dramatic increase in Rho signalling and consequent growth cone collapse, and this response is blocked by a dominant negative PDZRhoGEF (Perrot, Vazquez-Prado et al. 2002).

1.17. LARG and possibly PDZrhoGEF are activated by the insulin-like growth factor 1 receptor

LARG interacts with the insulin-like growth factor 1 (IGF-1) receptor via its PDZ domain (Taya, Inagaki et al. 2001). The IGF-1 receptor does not carry a canonical PDZ binding motif (see section 1.30) at its C-terminus (Gln-Ser-Ser-Thr-Cys-COOH). However, solely the C-terminal 20 amino acids were used in a yeast-2-hybrid assay that identified the interaction, so it does seem likely that the interaction is mediated via a classic C-terminus-PDZ domain interaction. The active IGF-1 receptor stimulates nucleotide exchange on RhoA in a cellular assay and this is inhibited by over-expressing solely the PDZ domain of LARG. IGF-1 stimulation also induces actin stress fibres, and this is partially inhibited by over-expression of the PDZ domain of LARG. Preliminary experiments indicate that PDZrhoGEF may also interact with the IGF-1 receptor (unpublished, (Taya, Inagaki et al. 2001)).

1.18. LARG and PDZrhoGEF are regulated by phosphorylation

In some cases Rho GEF activity is regulated by phosphorylation. For example, the Rac GEFs Vav1 and Vav2 are tightly regulated by tyrosine phosphorylation (Crespo, Schuebel et al. 1997). LARG is not able to stimulate guanine nucleotide exchange on Rho *in vitro* in direct response to $G\alpha_{12}$, although it can do so in a cellular system and in response to $G\alpha_{13}$ *in vitro*. This indicates the requirement for an extra factor in $G\alpha_{12}$ -mediated Rho activation by LARG. This factor is Tec, a tyrosine kinase, which phosphorylates LARG (but not p115RhoGEF) independently of $G\alpha_{12}$ activation and stimulates Rho signalling to a level similar to that achieved by $G\alpha_{13}$ in the absence of Tec (Suzuki, Nakamura et al. 2003). It is likely that the site of phosphorylation is within or nearby the RGS domain.

A separate study has shown an increase in Rho GEF activity in response to focal adhesion kinase (FAK)-mediated tyrosine phosphorylation of LARG and PDZrhoGEF (Chikumi, Fukuhara et al. 2002). In this case

the site of phosphorylation was determined to be C-terminal to the PH domain. PDZRhoGEF is also phosphorylated by p21-activated kinase (PAK) 4 C-terminally to the PH domain. This phosphorylation abrogates the ability of PDZRhoGEF to promote the formation of actin stress fibres in cells. The PAK family of serine-threonine kinases are regulated by Cdc42 and this could, therefore, represent a novel method of Cdc42 inhibition of Rho-mediated stress fibre formation.

These studies indicate that distinct signalling pathways may activate RGS Rho GEFs independently of GPCR-Gα pathways. Interestingly, the *DRhoGEF2* mutant phenotype during embryogenesis in *Drosophila* is stronger than that of *cta* or *fog* mutants, indicating the existence of Cta-independent signals feeding in to DRhoGEF2. The PDZ and C1 domains of DRhoGEF2 are also indicative of Cta-independent signalling through DRhoGEF2. Although there is no evidence of phosphorylation events regulating DRhoGEF2, it will be interesting to determine what these additional signals are, and how they regulate DRhoGEF2 function.

1.19. RGS Rho GEFs can homo- and heterodimerise via their C-termini

All three human RGS Rho GEFs are able to homodimerise. At present there is no evidence for DRhoGEF2 doing so, although it has not specifically been studied. PDZRhoGEF and LARG can also heterodimerise with each other, although p115RhoGEF can heterodimerise with neither (Chikumi, Barac et al. 2004). This dimerisation is mediated via a region C-terminal to the PH domain (**fig. 1.9**). This region is one of the least conserved amongst the RGS Rho GEFs and contains no obvious identifiable domains except a predicted coiled coil. Coiled coils are protein motifs that can mediate oligomerisation (Burkhard, Stetefeld et al. 2001).

Complete deletion of the C-terminus of any of the mammalian RGS Rho GEFs leads to an increase in Rho GEF activity *in vivo*. In the case of p115 Rho GEF/Lsc the mutation of residues within the predicted coiled coil abrogates homo-dimerisation although this has no effect on Rho GEF activity (Eisenhaure, Francis et al. 2003). This, and the fact that deletion of the C-terminus of p115RhoGEF does not increase its activity *in vitro*, indicates that there are likely to be other proteins interacting with the C-terminus and affecting Rho GEF activity *in vivo* (Wells, Gutowski et al. 2001). Since the C-terminus is the site of FAK phosphorylation on PDZRhoGEF and LARG, and serves to bind PDZRhoGEF to PAK4, it is possible that one of these interactors inhibits Rho GEF activity. For p115 Rho GEF C-terminal interactors have yet to be identified.

Isoforms of Lsc lacking its regulatory C-terminus are expressed in the mouse spleen (Eisenhaure, Francis et al. 2003). p115RhoGEF, PDZRhoGEF and LARG are focus-forming when over-expressed in human fibroblasts (Fukuhara, Chikumi et al. 2001) and deletion of the C-terminus significantly enhances their transforming potential (Chikumi, Barac et al. 2004). Again, it is possible that a factor interacting with the RGS Rho GEFs at the C-terminus serves to keep oncogenic potential in check, probably by inhibiting Rho GEF activity.

1.20. The Mutated in Colorectal Cancer (MCC) protein interacts with DRhoGEF2

There have been many binding partners identified for the human RGS Rho GEFs, but until recently none for DRhoGEF2. A yeast-2-hybrid screen (Fields and Song 1989) using the predicted PDZ domain of DRhoGEF2 yielded three potential binding partners (K. Barrett, unpublished). One of these, named *Drosophila* MCC (DMCC) due to its homology with human MCC, was subsequently shown to interact with recombinant DRhoGEF2 PDZ domain in co-immunoprecipitation experiments (K. Barrett, unpublished and here in chapter 5). Two of the human orthologues of DRhoGEF2, LARG and PDZRhoGEF, also co-

immunoprecipitate with the human orthologue of MCC (K.Barrett, unpublished). Thus, the interaction is conserved across species from fruit flies to humans. Both human and *Drosophila* MCC have a classic PDZ target motif at their C-terminus: Threonine - X – Leucine – COOH (where X is any amino acid). It is proposed, therefore, that MCC interacts with RGS Rho GEF via its C-terminus in a classic PDZ-target interaction (see section 1.30).

1.21. The MCC gene is mutated in a variety of human cancers

MCC was originally identified as a gene mutated in patients with colon cancer (Kinzler, Nilbert et al. 1991a). It was initially thought to be the tumour suppressor gene mutated in patients with familial adenomatous polyposis (FAP). This disease is characterised by the development of multiple pre-cancerous polyps (lesions) that have a 100% chance of becoming cancerous if the colon is not removed. However, the gene mutated in patients with FAP was subsequently found to be the adenomatous polyposis coli (APC) gene which lies within 500Kb of the MCC gene in the 5q21 region of the human genome (Kinzler, Nilbert et al. 1991b).

The evidence for MCC as a tumour suppressor is confusing. MCC is not the gene involved in FAP, but it does show frequent loss of heterozygosity in patients with colorectal cancer and in a number of other cancers including oral squamous cell carcinoma, gastric carcinoma and breast carcinoma (Medeiros, Nagai et al. 1994; Hsieh and Huang 1995; Huang, Chiang et al. 1997). The interpretation of this should be treated with caution, however, since when MCC is heterozygously deleted the remaining allele is not frequently mutated in colorectal tumours (Curtis, Bubb et al. 1994). This suggests that MCC does not function as a tumour suppressor in human colorectal cancer. However, evidence from knockout mice suggests that murine MCC does indeed function as a tumour suppressor. Knockout mice that are not expressing detectable MCC protein develop adenocarcinomas and tumours in other organs

including lung, liver and lymphoid tissue (Lipkin 1997). In a separate study, mouse lung tumours were assayed for MCC mRNA levels. Levels were decreased compared to normal lung tissue (Oreffo, Robinson et al. 1998). Overall, it is unclear whether MCC is acting as a tumour suppressor in humans but the available evidence, including the murine studies, indicates that there is a strong possibility that it is.

1.22. Murine MCC is expressed in intestinal epithelia and nervous tissue

An antibody directed against the C-terminal 12 amino acids of murine MCC indicates high levels of expression in the brain (Senda, Matsumine et al. 1999). Expression in cerebellar neurons is throughout the cell body, neuronal fibres and terminals, but no nuclear staining is observed. Closer examination using immunoelectron microscopy reveals an association of MCC with membranes, particularly organelle membranes such as mitochondria. It is interesting to note that MCC is expressed at high levels along with LARG and PDZrhoGEF in neurons of the mouse brain. It is tempting to speculate that this co-localisation could be indicative of a neuronal function for the interactors.

MCC expression is also observed in the intestinal epithelium, with strongest expression associated with the lateral membranes. Apical microvilli, hepatocytes of the liver and epithelial cells of the kidney are also strongly reactive to the antibody. In all cases of observed MCC expression it is mainly associating with a membrane, either plasma or organelle. Upon fractionation experiments of NIH3T3 fibroblast cells the MCC protein detectable in the crude membrane fraction is not solubilised with detergents implying that it is not a transmembrane protein, but is likely to be part of an insoluble complex (Matsumine, Senda et al. 1996).

1.23. MCC may play a role in cell cycle regulation

The over-expression of MCC in synchronous mouse fibroblasts in culture prevents them from entering S-phase (Matsumine, Senda et al. 1996).

Two mutant forms of MCC, one that was identified in a human colon cancer patient (Kinzler, Nilbert et al. 1991a), and another that disrupts a region with low homology to G protein-coupled muscarinic receptors, both abrogate the ability of MCC to block S-phase entry. This function may fulfil the tumour suppressor role that MCC is believed to possess, although there is no *in vivo* evidence for this as yet.

The same study indicated that MCC becomes phosphorylated upon serum stimulation of serum-starved cells. However, this phosphorylation is not cyclical in time with the cell cycle, so a role for MCC phosphorylation in regulation of the cell cycle is unlikely.

1.24. MCC has a homologue, MCC-2

MCC-2 was identified as an interactor of one of the PDZ domains of AIE-75, the antigen target in autoimmune enteropathy (AIE) (Ishikawa, Kobayashi et al. 2001). MCC-2 shares significant regions of identity with MCC including regions predicted to form coiled coils. The C-terminus of MCC-2 carries a classic PDZ target motif, Thr – X – Leu – COOH, as MCC does. However, only MCC-2 interacts with AIE-75 in immunoprecipitation assays. MCC-2 is expressed in skeletal muscle, liver, small intestine, placenta and lungs. Interestingly, it is not expressed in brain tissue, therefore MCC and MCC-2 have different expression patterns. Since MCC-2 and MCC do not share a binding partner in AIE-75 and they have different expression patterns, it is likely that they have different or partially over-lapping functions.

1.25. The interaction between MCC and RGS Rho GEFs may provide a link between Rho GTPases and cancer

Since Rho proteins regulate the cell cycle and migration they may be misregulated in the cellular progression to cancer. The interaction between MCC and PDZ Rho GEF/LARG could provide a link between Rho GTPases and cancer. The evidence linking Rho GTPase signalling pathways to primary human cancer is modest but growing. The vast

majority of studies of Rho GTPase signalling in cancer have involved the transformation of cells in culture. There have, so far, been few *in vivo* studies in animal models to assess the importance of Rho GTPase signalling in tumour growth and invasion, and this is one area where much more work needs to be done.

Unlike the Ras GTPases, Rho GTPases are not frequently mutated in tumours. One exception to this is the 5' untranslated region of the RhoH gene which is frequently rearranged in Non-Hodgkins lymphoma, and is a target of aberrant hypermutation activity in large-cell lymphomas. However, it is not known how these rearrangements and mutations affect RhoH activity (Ridley 2004).

The over-expression of Rho GTPases has been reported in a number of human cancers. Over-expression of a Rho GTPase leads to the same outcome as a constitutively active mutation, i.e. increased levels of signalling. Increased levels of Rho C can stimulate metastasis and RhoC is over-expressed in over 90% of cases of inflammatory breast cancer, a highly invasive cancer (Clark, Golub et al. 2000). Farnesyl transferase inhibitors (FTIs), which ultimately inhibit prenylation and therefore association of a GTPase with the membrane, may provide a therapy for RhoC over-expressing tumours. However, these inhibitors may function by inhibition of RhoB or RhoE rather than RhoC, since RhoC is not farnesylated (van Golen, Bao et al. 2002). Nevertheless, this indicates that over-expression is an important form of misregulation of Rho signalling in cancer.

RhoA function is required for another small GTPase, Ras, to transform cells and lead to anchorage-independent growth (Coleman, Marshall et al. 2004). This implicates RhoA as an oncogene, and is likely to relate to its effects on cell proliferation (see section 1.6.5). Conversely, RhoB is a gene with putative tumour suppressor properties. A recent study indicates that RhoB is able to inhibit phosphatidylinositol 3-kinase

signalling (Jiang, Sun et al. 2004). Since components of this pathway are frequently mutated in cancer this study may explain the tumour suppressor properties of RhoB.

Although the evidence for Rho GTPase involvement in cancer is limited, the evidence that Rho GEFs are misregulated in cancer is far more substantial. Dbl Rho GEF was originally identified as an oncogene in cell transformation assays (Eva and Aaronson 1985). Unlike Rho GTPases themselves, Rho GEFs do exhibit point mutations or deletions that lead to constitutive activation in human cancers.

LARG provides an example of a GEF that is misexpressed in cancer. This gene was discovered as an in-frame fusion with the mixed lineage leukaemia (MLL) gene in patients with acute myeloid leukaemia (Kourlas, Strout et al. 2000). This fusion gives rise to an incomplete LARG protein lacking its PDZ domain but containing the RGS, Dbl-homology and pleckstrin homology domains under the control of the MLL promoter.

Another GEF, T-cell lymphoma invasion and metastasis 1 (TIAM1), upregulates the expression of metalloprotease inhibitors and is likely to inhibit metastasis. Recently, a small molecule inhibitor of Rac that prevents its activation in response to TIAM1 has been tested in prostate cancer cells in culture (Gao, Dickerson et al. 2004). This inhibitor prevented the proliferation, anchorage-independent growth and invasion phenotypes that are associated with Rac1. The fact that the Rho GTPases and their GEFs are implicated in metastasis is not surprising given their primary function in regulating the actin cytoskeleton and cell migration.

Cancer cells proliferate unchecked and ultimately metastasise, but in order for a growing tumour to be sustainable, angiogenesis is required to maintain oxygen and nutrient supply. Recent work indicates a novel role for the semaphorin-plexin partnership in the promotion of angiogenesis. In an *in vivo* mouse model sema4D (a semaphorin ligand) potentially

enhances blood vessel formation, and this is plexin and Rho dependent (Basile, Barac et al. 2004). This opens a new area of study of Rho proteins in the regulation of angiogenesis in tumours.

Rho signalling is likely to become significant in the area of cancer therapeutics since there are already inhibitors of various Rho signalling pathway components available which may prove to have therapeutic uses. Examples include the farnesyl-transferase inhibitors mentioned above whose effects are rather non-specific, and small-molecule drugs targeting the PAK kinases and ROK kinases (Aznar, Fernandez-Valeron et al. 2004). Rigorous testing in animal models of these and other novel therapeutics will determine whether there is potential for Rho-targeted drugs in the clinic.

1.26. Two other proteins also interact with the PDZ domain of DRhoGEF2

The yeast-2-hybrid screen using the PDZ domain of DRhoGEF2 yielded two potential binding partners in addition to MCC. One has significant homology to human stomatin and C.elegans Mec-2, thus it is referred to as *DMec-2*. Mec-2 is required for the proper function of ion channels in neurons involved in the touch response (Goodman, Ernstrom et al. 2002). The third protein interacting with DRhoGEF2 PDZ domain, named *Group3*, does not have orthologues in species identifiable from Basic Local Alignment Search Tool (BLAST) searches (Altschul, Madden et al. 1997). This does not necessarily mean this gene has no function in *Drosophila* – for example the *grim*, *reaper* and *head involution defective* (*hid*) genes involved in apoptosis in *Drosophila* have functional mammalian orthologues with such limited sequence conservation that they are not picked up as orthologues in BLAST searches. The functions of Mec-2 and Group 3 proteins in *Drosophila* are under investigation.

All three proteins identified in the yeast 2-hybrid screen carry a classic PDZ target motif at their C-terminus: Thr – X – Leu – COOH. It is likely,

therefore, that they interact with the DRhoGEF2 PDZ domain via a classic PDZ domain-target interaction.

1.27. PDZ domains are protein-protein interaction modules

PDZ domains are ~90 amino acid protein modules that mediate protein interactions. Their name derives from some of the proteins in which they were initially discovered: Postsynaptic density 95, Discs large and Zonula occludens-1. Their structure consists of six β strands (β A- β F) and two α helices (α A and α B) which fold in an overall six-stranded β sandwich. They can be predicted from primary sequence data on the basis of conserved amino acid motifs and were originally termed GLGF domains due to the existence of a conserved Gly-Leu-Gly-Phe motif involved in binding the carboxylate group at the C-terminus of their target. PDZ (or PDZ-like) domains have been identified in organisms as diverse as bacteria, plants, yeast, fruit flies and humans (Ponting 1997). *Drosophila* is predicted to have 131 proteins containing PDZ domains based on predictions using Simple Modular Architecture Analysis Tool (SMART) (Letunic, Copley et al. 2004).

1.28. PDZ-containing proteins act as scaffolds and participate in signalling

In many cases proteins contain more than one PDZ domain and/or another type of protein interaction domain (such as Src homology 3 (SH3)) in addition to their PDZ domain, and for this reason they are often thought of as “scaffold” proteins that function to bring together signalling proteins. One protein, MUPP1, contains no less than 13 PDZ domains (Ullmer, Schmuck et al. 1998). In *Drosophila* the signalling protein InaD, which contains 5 PDZ domains, acts as a scaffold for proteins involved in photoreceptor signal transduction (Xu, Choudhury et al. 1998). The function of scaffolds in providing close physical proximity of the signalling complex machinery is likely to allow the rapid processing of signals along the pathway and tight regulation.

Although PDZ domain-containing proteins may solely function as a scaffold, they often play more than a passive role, participating in the transduction of signals via enzymatic or other domains which they also often contain. LIM kinase is such an example with its serine-threonine kinase domain in addition to its PDZ domain.

The range of functions provided by PDZ domain-containing proteins is vast, not surprisingly given their common multi-domain structure. All characterised thus far are cytoplasmic (except interleukin-16), and many are associated with the plasma membrane. PDZ domain-containing proteins can generally be grouped into one of three sub-categories; those involved in signalling pathways as adaptors for cell surface receptors (including receptor tyrosine kinases and GPCRs), those involved in the establishment of epithelial polarity or those involved in signalling at both the pre- and post-synaptic sites of neurons (Nourry, Grant et al. 2003).

1.29. There are several mechanisms of interaction between PDZ domains and their targets

The canonical method of interaction between PDZ domain and target is via the C-terminus of the target. In this case the terminal amino acids insert into a “pocket” created by residues of the α B helix and β B strand (fig. 1.10). In a second, less common arrangement, the PDZ domain binds to an internal sequence of the target protein. For example, the transient receptor potential Ca^{2+} channel contains an internal PDZ binding motif that interacts with one of the InaD PDZ domains (Chevesich, Kreuz et al. 1997). PDZ domains can also dimerise, either with themselves (homo-) or with another PDZ domain (hetero-). An example of heterodimerisation involves amino acid residues of the nNOS PDZ domain forming a two-stranded β “finger” which inserts into the psd-95 PDZ binding pocket (Tochio, Mok et al. 2000). Interestingly, the sixth PDZ domain of GRIP mediates homodimerisation but the crystal structure indicates that the peptide binding groove is not involved indicating that

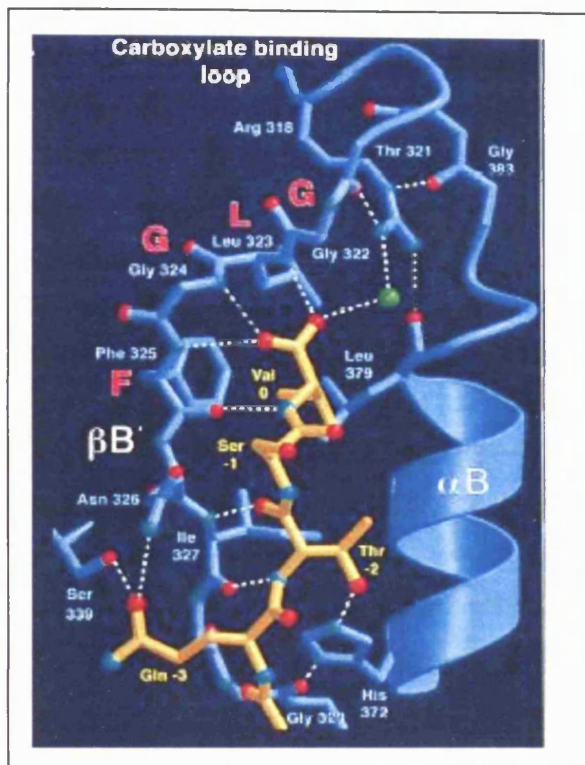


Figure 1.10. The PDZ-target interaction.

Representation of the crystal structure of part of psd95-PDZ3 (blue) in complex with peptide (yellow). Taken from (Doyle, Lee et al. 1996).

targets could simultaneously bind, and a molecular complex be built up around the interaction (Im, Park et al. 2003).

Recently, a group of class II PDZ domains (see table 1.2) have also been shown to mediate interaction with phosphatidylinositol-3,4-diphosphate (PIP₂), a plasma membrane lipid (Zimmermann, Meerschaert et al. 2002). Interestingly, excess PIP₂ interferes with the PDZ:target peptide interaction indicating that perhaps the lipid binding site overlaps with the peptide binding site. To

date, only a small group of class II PDZ domains have been shown to interact with lipids in this way, but perhaps other classes of PDZ domains can do the same.

1.30. The specificity of the PDZ-target interaction is determined by amino acid residues at the C-terminus of the target

The crystal structures of the Psd-95 PDZ domain complexed with its target indicate that the C-terminal 4 amino acids form specific interactions with the peptide-binding groove (fig. 1.10) (Doyle, Lee et al. 1996). The residue at the 0 position (i.e. the most C-terminal amino acid) is always hydrophobic in nature (e.g. leucine, valine, isoleucine), and this inserts into a hydrophobic pocket. In many cases, the residue at the -1 position

points away from the interaction surface and thus does not participate. In this respect, the -1 position is highly variable amongst PDZ targets. However, there are examples of the -1 position contributing directly to the interaction e.g. a cysteine at the -1 position of NorpA plays a crucial role in its interaction with the first PDZ domain of InaD (Kimple, Siderovski et al. 2001).

The amino acid at the -2 position is important for binding and is used to classify PDZ domains based on their -2 position binding specificity (**table 1.2**) (Daniels, Cohen et al. 1998). Class I PDZ domains have a histidine residue at position $\alpha B1$ (the bottom of the αB helix, **fig. 1.10**) which forms a hydrogen bond with the -2 position serine or threonine residue of the target. The vast majority of PDZ domains classified thus far fall into this class. Class II PDZ domains, by contrast, have a hydrophobic amino acid at the $\alpha B1$ position which forms an interaction with a hydrophobic amino acid at the -2 position of the target. The final class, III, prefer negatively charged amino acids at the -2 position of the target.

It should be noted that some PDZ domains can bind multiple targets that place them in more than one class (Nourry, Grant et al. 2003).

1.31. The PDZ domain of DRhoGEF2 is a class I PDZ domain

In light of the fact that all three targets picked up in the yeast-2-hybrid screen with DRhoGEF2 PDZ domain have Thr-X-Leu-COOH at their C-terminus, it would seem that DRhoGEF2 is a class I PDZ domain. It does indeed have a histidine residue at the $\alpha B1$ position, indicative of a class I PDZ domain (Songyang, Fanning et al. 1997).

1.32. Amino acids N-terminal to the -2 position are also important in the specificity of binding

Although there is promiscuity between PDZ domains and their targets, it is likely that some degree of specificity greater than the class system determines which PDZ domain protein can bind which target.

Class	C-terminal sequence	Interacting protein	PDZ domain-containing protein
Class I			
X-S or T-X-V	E-S-D-V	NMDAR2A, B	PSD-95 (PDZ2)
	E-T-D-V	Shaker-type K ⁺ channel	
	D-S-W-V	p0071	Erbin
		δ -catenin, ARVCF	
X-S or T-X-L	D-S-S-L	β_2 -adrenergic receptor	NHERF or EBP50
	Q-T-R-L	GKAP	Shank pr ProSAP
Class II			
X- Ψ -X- Ψ	E-Y-Y-V	Neurexin	CASK
	E-F-Y-A	Syndecan	CASK, syntenin
	S-V-E-V	EphB2	PICK1
	D-V-P-V	ErbB2	Erbin
Class III			
X-D or E-X- Ψ	V-D-S-V	Melatonin receptor	nNOS
	G-E-P-L	KIF17	mLIN10 or Mint1 or X11

Table 1.2. The classification of PDZ domains. PDZ domains are classified based on the C-terminal motif to which they bind. Taken from (Nourry, Grant et al. 2003).

Since the group of class I PDZ domains is large (at least 70 across species), and since more than one PDZ-containing protein is present in a single cell, it is likely that there are further means of regulating specificity of binding. Subcellular localisation could play a major role in determining whether a PDZ domain interacts with one target or another, but the fact that most PDZ domain proteins are associated with the plasma membrane indicates that there are likely to be further modes of regulation.

An elegant study using a peptide library approach determined that PDZ domains were generally selective for up to 7 amino acid residues from the C-terminus (i.e. up to the -6 position) (Songyang, Fanning et al. 1997). The solution structure of the second PDZ domain of human phosphatase hPTP1E indicates that target residues up to the -5 position are involved in binding the PDZ domain (Kozlov, Banville et al. 2002). In this case, the

β B- β C loop mediates interaction with the amino acid at the -5 position providing a new region of focus in the determination of PDZ-target specificity. A synthetic peptide library screening approach with this same PDZ domain suggested that up to the -8 position may be involved in binding (Songyang, Fanning et al. 1997).

1.33. PDZ domains can be engineered to bind a specific C-terminal target sequence

An elegant mutagenesis screen using a yeast-2-hybrid approach identified novel PDZ domain sequences that can bind specific C-terminal target peptides (Schneider, Buchert et al. 1999). These were then verified in a mammalian cell system. In two cases, orphan PDZ targets were used as bait for mutagenised PDZ domains in order to identify specific residues required for binding. Most residues identified were in the carboxylate binding loop, the β B strand or α B helix (fig. 1.10). However, some residues were in regions outside the binding groove indicating that these can contribute to specificity.

A second approach used computational predictions based on known structural information to mutate a PDZ domain from class I to class II (Reina, Lacroix et al. 2002). In this case the specificity of interactions was verified in yeast-2-hybrid experiments, "affinity purification" using an immobilised PDZ domain, and in one case a protein pull-down. When the PDZ domain was mutated to bind a new target that differed only in the -1 and -3 positions (i.e. was of the same class), no less than 6 amino acid changes were predicted. In order to turn a class I binder into a class II binder, 7 amino acid changes were made, the most critical being the mutation of the α B1 position histidine residue into a leucine in order to accommodate the hydrophobic amino acid at the -2 position of the target. Similarly to the findings by Schneider *et al*, the vast majority of mutations affected the carboxylate binding loop, the β B strand or α B helix.

In a further study, three different PDZ domains were tested for their ability to bind library peptides and showed differing degrees of promiscuity towards target sequences (Wiedemann, Boisguerin et al. 2004). The SNA1 (α -1 syntrophin) PDZ domain recognises a much larger set of C-terminal sequences than the ERBIN (ErbB2 interacting protein) PDZ domain. This may reflect their different biological functions *in vivo* since SNA1 is a scaffolding protein whose function is likely to be in the construction of large protein complexes, whereas ERBIN is a suppressor of the Ras/Raf pathway.

These studies are interesting in determining the critical residues required for binding, and are beginning to map out regions that determine specificity of binding. However, there is still a long way to go before the specificity of binding *in vivo* is understood.

The PDZ-mediated interaction between MCC and RGS RhoGEF forms the basis of study for this thesis. Since the interaction appears to be conserved from flies to humans it is likely to be of significance to organismal development and survival.

1.34. The hypothesis for this thesis

The work in this thesis is based on the following hypothesis:

The Drosophila melanogaster orthologue of human MCC interacts with DRhoGEF2 during development, and this interaction affects DRhoGEF2 signalling and Drosophila development.

MCC is a potentially interesting protein to study due to its predicted tumour suppressor function in humans. There are no known interactors for MCC1 in humans, and its exact mechanism of action as a tumour suppressor is unknown. DRhoGEF2 interactions are interesting to analyse since *DRhoGEF2* mutants have a stronger phenotype than *cta* or *fog* mutants. This indicates that there may be other signalling pathways

feeding in to or out of DRhoGEF2, in which DMCC (*Drosophila melanogaster* MCC) could participate. Preliminary study of human MCC in mammalian tissue culture cells indicates that it may affect the actin cytoskeleton (K. Barrett, unpublished data), which could imply a link with Rho signalling. These factors make the interaction between DRhoGEF2 and DMCC an interesting one to study.

Drosophila melanogaster was chosen as a model organism to study MCC and its interaction with RhoGEF due to the existence of only one MCC orthologue (thereby overcoming redundancy problems), and the ease of genetic and cell biology analysis.

1.35. The aims of this study

In order to prove or disprove the above hypothesis, this study looks to achieve two main aims. The first is to characterise the function of MCC in *Drosophila* development, and the second is to better understand the relationship between DRhoGEF2 and DMCC. The work can be broadly grouped into three areas:

- 1) MCC gene and protein analysis (chapter 3)
- 2) Analysis of MCC loss- and gain-of-function (chapter 4)
- 3) Analysis of MCC interaction with DRhoGEF2 (chapter 5)

1.36. Chapter 1 References

- Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of *Drosophila melanogaster*." Science **287**(5461): 2185-95.
- Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.
- Amano, M., M. Ito, et al. (1996). "Phosphorylation and activation of myosin by Rho-associated kinase (Rho-kinase)." J Biol Chem **271**(34): 20246-9.

- Aurandt, J., H. G. Vikis, et al. (2002). "The semaphorin receptor plexin-B1 signals through a direct interaction with the Rho-specific nucleotide exchange factor, LARG." Proc Natl Acad Sci U S A **99**(19): 12085-90.
- Aznar, S., P. Fernandez-Valeron, et al. (2004). "Rho GTPases: potential candidates for anticancer therapy." Cancer Lett **206**(2): 181-91.
- Banerjee, J. and P. B. Wedegaertner (2004). "Identification of a novel sequence in PDZ-RhoGEF that mediates interaction with the actin cytoskeleton." Mol Biol Cell **15**(4): 1760-75.
- Barrett, K., M. Leptin, et al. (1997). "The Rho GTPase and a putative RhoGEF mediate a signaling pathway for the cell shape changes in *Drosophila* gastrulation." Cell **91**(7): 905-15.
- Basile, J. R., A. Barac, et al. (2004). "Class IV semaphorins promote angiogenesis by stimulating Rho-initiated pathways through plexin-B." Cancer Res **64**(15): 5212-24.
- Becknell, B., T. Shen, et al. (2003). "Characterization of leukemia-associated Rho guanine nucleotide exchange factor (LARG) expression during murine development." Cell Tissue Res **314**(3): 361-6.
- Bertet, C., L. Sulak, et al. (2004). "Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation." Nature **429**(6992): 667-71.
- Bishop, A. L. and A. Hall (2000). "Rho GTPases and their effector proteins." Biochem J **348 Pt 2**: 241-55.
- Bloor, J. W. and D. P. Kiehart (2002). "Drosophila RhoA regulates the cytoskeleton and cell-cell adhesion in the developing epidermis." Development **129**(13): 3173-83.
- Booden, M. A., D. P. Siderovski, et al. (2002). "Leukemia-associated Rho guanine nucleotide exchange factor promotes G alpha q-coupled activation of RhoA." Mol Cell Biol **22**(12): 4053-61.
- Braga, V. (2000). "Epithelial cell shape: cadherins and small GTPases." Exp Cell Res **261**(1): 83-90.

- Braga, V. M. (2002). "GEF without a Dbl domain?" Nat Cell Biol **4**(8): E188-90.
- Burkhard, P., J. Stetefeld, et al. (2001). "Coiled coils: a highly versatile protein folding motif." Trends Cell Biol **11**(2): 82-8.
- Burridge, K. and K. Wennerberg (2004). "Rho and rac take center stage." Cell **116**(2): 167-79.
- Caron, E. and A. Hall (1998). "Identification of two distinct mechanisms of phagocytosis controlled by different Rho GTPases." Science **282**(5394): 1717-21.
- Chang, H. Y. and D. F. Ready (2000). "Rescue of photoreceptor degeneration in rhodopsin-null Drosophila mutants by activated Rac1." Science **290**(5498): 1978-80.
- Chen, G. C., P. Gajowniczek, et al. (2004). "Rho-LIM kinase signaling regulates ecdysone-induced gene expression and morphogenesis during Drosophila metamorphosis." Curr Biol **14**(4): 309-13.
- Cheng, L., K. L. Rossman, et al. (2002). "RhoGEF specificity mutants implicate RhoA as a target for Dbs transforming activity." Mol Cell Biol **22**(19): 6895-905.
- Chevesich, J., A. J. Kreuz, et al. (1997). "Requirement for the PDZ domain protein, INAD, for localization of the TRP store-operated channel to a signaling complex." Neuron **18**(1): 95-105.
- Chikumi, H., A. Barac, et al. (2004). "Homo- and hetero-oligomerization of PDZ-RhoGEF, LARG and p115RhoGEF by their C-terminal region regulates their in vivo Rho GEF activity and transforming potential." Oncogene **23**(1): 233-40.
- Chikumi, H., S. Fukuhara, et al. (2002). "Regulation of G protein-linked guanine nucleotide exchange factors for Rho, PDZ-RhoGEF, and LARG by tyrosine phosphorylation: evidence of a role for focal adhesion kinase." J Biol Chem **277**(14): 12463-73.
- Clark, E. A., T. R. Golub, et al. (2000). "Genomic analysis of metastasis reveals an essential role for RhoC." Nature **406**(6795): 532-5.

- Coleman, M. L., C. J. Marshall, et al. (2004). "RAS and RHO GTPases in G1-phase cell-cycle regulation." Nat Rev Mol Cell Biol 5(5): 355-66.
- Conway, A. M., A. B. James, et al. (2004). "Regulation of myosin light chain phosphorylation by RhoB in neuronal cells." Exp Cell Res 300(1): 35-42.
- Costa, M., E. T. Wilson, et al. (1994). "A putative cell signal encoded by the folded gastrulation gene coordinates cell shape changes during Drosophila gastrulation." Cell 76: 1075-1089.
- Crawford, J. M., N. Harden, et al. (1998). "Cellularization in Drosophila melanogaster is disrupted by the inhibition of rho activity and the activation of Cdc42 function." Dev Biol 204(1): 151-64.
- Crespo, P., K. E. Schuebel, et al. (1997). "Phosphotyrosine-dependent activation of Rac-1 GDP/GTP exchange by the vav proto-oncogene product." Nature 385(6612): 169-72.
- Curtis, L. J., V. J. Bubbs, et al. (1994). "Loss of heterozygosity of MCC is not associated with mutation of the retained allele in sporadic colorectal cancer." Hum Mol Genet 3(3): 443-6.
- Daniels, D. L., A. R. Cohen, et al. (1998). "Crystal structure of the hCASK PDZ domain reveals the structural basis of class II PDZ domain target recognition." Nat Struct Biol 5(4): 317-25.
- Daub, H., K. Gevaert, et al. (2001). "Rac/Cdc42 and p65PAK regulate the microtubule-destabilizing protein stathmin through phosphorylation at serine 16." J Biol Chem 276(3): 1677-80.
- Doyle, D. A., A. Lee, et al. (1996). "Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ." Cell 85(7): 1067-76.
- Duchek, P. and P. Rorth (2001). "Guidance of cell migration by EGF receptor signaling during Drosophila oogenesis." Science 291(5501): 131-3.
- Eaton, S., P. Auvinen, et al. (1995). "CDC42 and Rac1 control different actin-dependent processes in the Drosophila wing disc epithelium." J Cell Biol 131(1): 151-64.

- Eisenhaure, T. M., S. A. Francis, et al. (2003). "The Rho guanine nucleotide exchange factor Lsc homo-oligomerizes and is negatively regulated through domains in its carboxyl terminus that are absent in novel splenic isoforms." J Biol Chem **278**(33): 30975-84.
- Erickson, J. W. and R. A. Cerione (2004). "Structural elements, mechanism, and evolutionary convergence of Rho protein-guanine nucleotide exchange factor complexes." Biochemistry **43**(4): 837-42.
- Etienne-Manneville, S. and A. Hall (2001). "Integrin-mediated activation of Cdc42 controls cell polarity in migrating astrocytes through PKCzeta." Cell **106**(4): 489-98.
- Etienne-Manneville, S. and A. Hall (2002). "Rho GTPases in cell biology." Nature **420**(6916): 629-35.
- Etienne-Manneville, S. and A. Hall (2003). "Cdc42 regulates GSK-3beta and adenomatous polyposis coli to control cell polarity." Nature **421**(6924): 753-6.
- Eva, A. and S. A. Aaronson (1985). "Isolation of a new human oncogene from a diffuse B-cell lymphoma." Nature **316**(6025): 273-5.
- Fanto, M., U. Weber, et al. (2000). "Nuclear signaling by Rac and Rho GTPases is required in the establishment of epithelial planar polarity in the Drosophila eye." Curr Biol **10**(16): 979-88.
- Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-6.
- Fuentes, E. J., A. E. Karnoub, et al. (2003). "Critical role of the pleckstrin homology domain in Dbs signaling and growth regulation." J Biol Chem **278**(23): 21188-96.
- Fukuhara, S., H. Chikumi, et al. (2001). "RGS-containing RhoGEFs: the missing link between transforming G proteins and Rho?" Oncogene **20**(13): 1661-8.
- Fukuhara, S., C. Murga, et al. (1999). "A novel PDZ domain containing guanine nucleotide exchange factor links heterotrimeric G proteins to Rho." J Biol Chem **274**(9): 5868-79.

- Gao, Y., J. B. Dickerson, et al. (2004). "Rational design and characterization of a Rac GTPase-specific small molecule inhibitor." Proc Natl Acad Sci U S A **101**(20): 7618-23.
- Genova, J. L., S. Jong, et al. (2000). "Functional analysis of Cdc42 in actin filament assembly, epithelial morphogenesis, and cell signaling during *Drosophila* development." Dev Biol **221**(1): 181-94.
- Girkontaite, I., K. Missy, et al. (2001). "Lsc is required for marginal zone B cells, regulation of lymphocyte motility and immune responses." Nat Immunol **2**(9): 855-62.
- Glotzer, M. (2001). "Animal cell cytokinesis." Annu Rev Cell Dev Biol **17**: 351-86.
- Goodman, M. B., G. G. Ernst, et al. (2002). "MEC-2 regulates *C. elegans* DEG/ENAC channels needed for mechanosensation." Nature **415**(6875): 1039-42.
- Grosshans, J., C. Wenzl, et al. (2005). "RhoGEF2 and the formin Dia control the formation of the furrow canal by directed actin assembly during *Drosophila* cellularisation." Development **132**(5): 1009-1020.
- Grosshans, J. and E. Wieschaus (2000). "A genetic link between morphogenesis and cell division during formation of the ventral furrow in *Drosophila*." Cell **101**(5): 523-31.
- Hacker, U. and N. Perrimon (1998). "DRhoGEF2 encodes a member of the Dbl family of oncogenes and controls cell shape changes during gastrulation in *Drosophila*." Genes Dev **12**(2): 274-84.
- Hakeda-Suzuki, S., J. Ng, et al. (2002). "Rac function and regulation during *Drosophila* development." Nature **416**(6879): 438-42.
- Harden, N., M. Ricos, et al. (2002). "Drac1 and Crumbs participate in amnioserosa morphogenesis during dorsal closure in *Drosophila*." J Cell Sci **115**(Pt 10): 2119-29.
- Hariharan, I. K., K. Q. Hu, et al. (1995). "Characterization of rho GTPase family homologues in *Drosophila melanogaster*: overexpressing

- Rho1 in retinal cells causes a late developmental defect." Embo J **14**(2): 292-302.
- Hart, M. J., X. Jiang, et al. (1998). "Direct stimulation of the guanine nucleotide exchange activity of p115 RhoGEF by G α 13." Science **280**(5372): 2112-4.
- Hart, M. J., S. Sharma, et al. (1996). "Identification of a novel guanine nucleotide exchange factor for the Rho GTPase." J Biol Chem **271**(41): 25452-8.
- Hsieh, L. L. and Y. C. Huang (1995). "Loss of heterozygosity of APC/MCC gene in differentiated and undifferentiated gastric carcinomas in Taiwan." Cancer Lett **96**(2): 169-74.
- Huang, J. S., C. P. Chiang, et al. (1997). "Loss of heterozygosity of APC and MCC genes in oral squamous cell carcinomas in Taiwan." J Oral Pathol Med **26**(7): 322-6.
- Hutson, M. S., Y. Tokutake, et al. (2003). "Forces for morphogenesis investigated with laser microsurgery and quantitative modeling." Science **300**(5616): 145-9.
- Im, Y. J., S. H. Park, et al. (2003). "Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization." J Biol Chem **278**(10): 8501-7.
- Ishikawa, S., I. Kobayashi, et al. (2001). "Interaction of MCC2, a novel homologue of MCC tumor suppressor, with PDZ-domain Protein AIE-75." Gene **267**(1): 101-10.
- Jacinto, A., W. Wood, et al. (2000). "Dynamic actin-based epithelial adhesion and cell matching during Drosophila dorsal closure." Curr Biol **10**(22): 1420-6.
- Jacinto, A., W. Wood, et al. (2002). "Dynamic analysis of actin cable function during Drosophila dorsal closure." Curr Biol **12**(14): 1245-50.
- Jackson, M., W. Song, et al. (2001). "Modulation of the neuronal glutamate transporter EAAT4 by two interacting proteins." Nature **410**(6824): 89-93.

- Jiang, K., J. Sun, et al. (2004). "Akt mediates Ras downregulation of RhoB, a suppressor of transformation, invasion, and metastasis." Mol Cell Biol **24**(12): 5565-76.
- Karnoub, A. E., M. Symons, et al. (2004). "Molecular basis for Rho GTPase signaling specificity." Breast Cancer Res Treat **84**(1): 61-71.
- Karnoub, A. E., D. K. Worthylake, et al. (2001). "Molecular basis for Rac1 recognition by guanine nucleotide exchange factors." Nat Struct Biol **8**(12): 1037-41.
- Kimple, M. E., D. P. Siderovski, et al. (2001). "Functional relevance of the disulfide-linked complex of the N-terminal PDZ domain of InaD with NorpA." Embo J **20**(16): 4414-22.
- Kinzler, K. W., M. C. Nilbert, et al. (1991). "Identification of FAP locus genes from chromosome 5q21." Science **253**(5020): 661-5.
- Kinzler, K. W., M. C. Nilbert, et al. (1991). "Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers." Science **251**(4999): 1366-70.
- Kolega, J. (2003). "Asymmetric distribution of myosin IIB in migrating endothelial cells is regulated by a rho-dependent kinase and contributes to tail retraction." Mol Biol Cell **14**(12): 4745-57.
- Kourlas, P. J., M. P. Strout, et al. (2000). "Identification of a gene at 11q23 encoding a guanine nucleotide exchange factor: evidence for its fusion with MLL in acute myeloid leukemia." Proc Natl Acad Sci U S A **97**(5): 2145-50.
- Kozasa, T., X. Jiang, et al. (1998). "p115 RhoGEF, a GTPase activating protein for Galpha12 and Galpha13." Science **280**(5372): 2109-11.
- Kozlov, G., D. Banville, et al. (2002). "Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the beta2-beta3 loop to PDZ domain-ligand interactions." J Mol Biol **320**(4): 813-20.
- Kozma, R., S. Ahmed, et al. (1995). "The Ras-related protein Cdc42Hs and bradykinin promote formation of peripheral actin microspikes

- and filopodia in Swiss 3T3 fibroblasts." Mol Cell Biol **15**(4): 1942-52.
- Kranenburg, O., M. Poland, et al. (1999). "Activation of RhoA by lysophosphatidic acid and G α 12/13 subunits in neuronal cells: induction of neurite retraction." Mol Biol Cell **10**(6): 1851-7.
- Kuner, R., J. M. Swiercz, et al. (2002). "Characterization of the expression of PDZ-RhoGEF, LARG and G α 12/G α 13 proteins in the murine nervous system." Eur J Neurosci **16**(12): 2333-41.
- Lee, T., C. Winter, et al. (2000). "Essential roles of Drosophila RhoA in the regulation of neuroblast proliferation and dendritic but not axonal morphogenesis." Neuron **25**(2): 307-16.
- Letunic, I., R. R. Copley, et al. (2004). "SMART 4.0: towards genomic data integration." Nucleic Acids Res **32 Database issue**: D142-4.
- Lin, D., A. S. Edwards, et al. (2000). "A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity." Nat Cell Biol **2**(8): 540-7.
- Lipkin, M. (1997). "New rodent models for studies of chemopreventive agents." J Cell Biochem Suppl **28-29**: 144-7.
- Luna, A., O. B. Matas, et al. (2002). "Regulation of protein transport from the Golgi complex to the endoplasmic reticulum by CDC42 and N-WASP." Mol Biol Cell **13**(3): 866-79.
- Lundstrom, A., M. Gallio, et al. (2004). "Vilse, a conserved Rac/Cdc42 GAP mediating Robo repulsion in tracheal cells and axons." Genes Dev **18**(17): 2161-71.
- Magie, C. R., M. R. Meyer, et al. (1999). "Mutations in the Rho1 small GTPase disrupt morphogenesis and segmentation during early Drosophila development." Development **126**(23): 5353-64.
- Marlow, F., J. Topczewski, et al. (2002). "Zebrafish Rho kinase 2 acts downstream of Wnt11 to mediate cell polarity and effective convergence and extension movements." Curr Biol **12**(11): 876-84.

- Matsumine, A., T. Senda, et al. (1996). "MCC, a cytoplasmic protein that blocks cell cycle progression from the G0/G1 to S phase." J Biol Chem **271**(17): 10341-6.
- Medeiros, A. C., M. A. Nagai, et al. (1994). "Loss of heterozygosity affecting the APC and MCC genetic loci in patients with primary breast carcinomas." Cancer Epidemiol Biomarkers Prev **3**(4): 331-3.
- Meller, N., M. Irani-Tehrani, et al. (2004). "The novel Cdc42 guanine nucleotide exchange factor, zizimin1, dimerizes via the Cdc42-binding CZH2 domain." J Biol Chem **279**(36): 37470-6.
- Murphy, A. M. and D. J. Montell (1996). "Cell type-specific roles for Cdc42, Rac, and RhoL in Drosophila oogenesis." J Cell Biol **133**(3): 617-30.
- Ng, J., T. Nardine, et al. (2002). "Rac GTPases control axon growth, guidance and branching." Nature **416**(6879): 442-7.
- Nikolaidou, K. K. and K. Barrett (2004). "A Rho GTPase Signaling Pathway Is Used Reiteratively in Epithelial Folding and Potentially Selects the Outcome of Rho Activation." Curr Biol **14**(20): 1822-6.
- Nobes, C. D. and A. Hall (1995). "Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia." Cell **81**(1): 53-62.
- Nolan, K. M., K. Barrett, et al. (1998). "Myoblast city, the Drosophila homolog of DOCK180/CED-5, is required in a Rac signaling pathway utilized for multiple developmental processes." Genes Dev **12**(21): 3337-42.
- Nourry, C., S. G. Grant, et al. (2003). "PDZ domain proteins: plug and play!" Sci STKE **2003**(179): RE7.
- Oinuma, I., Y. Ishikawa, et al. (2004). "The Semaphorin 4D receptor Plexin-B1 is a GTPase activating protein for R-Ras." Science **305**(5685): 862-5.
- Oleksy, A., H. Barton, et al. (2004). "Preliminary crystallographic analysis of the complex of the human GTPase RhoA with the DH/PH

- tandem of PDZ-RhoGEF." Acta Crystallogr D Biol Crystallogr **60**(Pt 4): 740-2.
- Olson, M. F., A. Ashworth, et al. (1995). "An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1." Science **269**(5228): 1270-2.
- Oreffo, V. I., S. Robinson, et al. (1998). "Decreased expression of the adenomatous polyposis coli (Apc) and mutated in colorectal cancer (Mcc) genes in mouse lung neoplasia." Mol Carcinog **21**(1): 37-49.
- Padash Barmchi, M., S. Rogers, et al. (2005). "DRhoGEF2 regulates actin organization and contractility in the Drosophila blastoderm embryo." J Cell Biol.
- Palazzo, A. F., C. H. Eng, et al. (2004). "Localized stabilization of microtubules by integrin- and FAK-facilitated Rho signaling." Science **303**(5659): 836-9.
- Parks, S. and E. Wieschaus (1991). "The Drosophila gastrulation gene *concertina* encodes a G alpha-like protein." Cell **64**: 447-458.
- Perrot, V., J. Vazquez-Prado, et al. (2002). "Plexin B regulates Rho through the guanine nucleotide exchange factors Leukemia-associated RhoGEF (LARG) and PDZ-RhoGEF." J Biol Chem **14**: 14.
- Ponting, C. P. (1997). "Evidence for PDZ domains in bacteria, yeast, and plants." Protein Sci **6**(2): 464-8.
- Pruyne, D. and A. Bretscher (2000). "Polarization of cell growth in yeast. I. Establishment and maintenance of polarity states." J Cell Sci **113** (Pt 3): 365-75.
- Qualmann, B. and H. Mellor (2003). "Regulation of endocytic traffic by Rho GTPases." Biochem J **371**(Pt 2): 233-41.
- Reina, J., E. Lacroix, et al. (2002). "Computer-aided design of a PDZ domain to recognize new target sequences." Nat Struct Biol **9**(8): 621-7.
- Ridley, A. J. (2004). "Rho proteins and cancer." Breast Cancer Res Treat **84**(1): 13-9.

- Ridley, A. J. and A. Hall (1992). "The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors." Cell **70**(3): 389-99.
- Ridley, A. J., H. F. Paterson, et al. (1992). "The small GTP-binding protein rac regulates growth factor-induced membrane ruffling." Cell **70**(3): 401-10.
- Rivero, F., H. Dislich, et al. (2001). "The Dictyostelium discoideum family of Rho-related proteins." Nucleic Acids Res **29**(5): 1068-79.
- Rogers, S. L., U. Wiedemann, et al. (2004). "Drosophila RhoGEF2 associates with microtubule plus ends in an EB1-dependent manner." Curr Biol **14**(20): 1827-33.
- Rumenapp, U., A. Blomquist, et al. (1999). "Rho-specific binding and guanine nucleotide exchange catalysis by KIAA0380, a dbf family member." FEBS Lett **459**(3): 313-8.
- Rumenapp, U., A. Blomquist, et al. (1999). "Rho-specific binding and guanine nucleotide exchange catalysis by KIAA0380, a dbf family member." FEBS Lett **459**(3): 313-8.
- Schneider, S., M. Buchert, et al. (1999). "Mutagenesis and selection of PDZ domains that bind new protein targets." Nat Biotechnol **17**(2): 170-5.
- Schumacher, S., T. Gryzik, et al. (2004). "The RhoGEF Pebble is required for cell shape changes during cell migration triggered by the Drosophila FGF receptor Heartless." Development **131**(11): 2631-40.
- Senda, T., A. Matsumine, et al. (1999). "Localization of MCC (mutated in colorectal cancer) in various tissues of mice and its involvement in cell differentiation." J Histochem Cytochem **47**(9): 1149-58.
- Sepp, K. J. and V. J. Auld (2003). "RhoA and Rac1 GTPases mediate the dynamic rearrangement of actin in peripheral glia." Development **130**(9): 1825-35.
- Settleman, J., Barrett, K. (2001). "Genetic analysis of Rho GTPase function." Recent Res. Devel. Mol. Cell. Biol. **2**: 105-123.

-
- Shandala, T., S. L. Gregory, et al. (2004). "Citron Kinase is an essential effector of the Pbl-activated Rho signalling pathway in *Drosophila melanogaster*." Development **131**(20): 5053-63.
- Small, J. V., T. Stradal, et al. (2002). "The lamellipodium: where motility begins." Trends Cell Biol **12**(3): 112-20.
- Snyder, J. T., D. K. Worthylake, et al. (2002). "Structural basis for the selective activation of Rho GTPases by Dbl exchange factors." Nat Struct Biol **9**(6): 468-75.
- Sone, M., M. Hoshino, et al. (1997). "Still life, a protein in synaptic terminals of *Drosophila* homologous to GDP-GTP exchangers." Science **275**(5299): 543-7.
- Songyang, Z., A. S. Fanning, et al. (1997). "Recognition of unique carboxyl-terminal motifs by distinct PDZ domains." Science **275**(5296): 73-7.
- Speck, O., S. C. Hughes, et al. (2003). "Moesin functions antagonistically to the Rho pathway to maintain epithelial integrity." Nature **421**(6918): 83-7.
- Strutt, D. I., U. Weber, et al. (1997). "The role of RhoA in tissue polarity and Frizzled signalling." Nature **387**(6630): 292-5.
- Suzuki, N., S. Nakamura, et al. (2003). "Galpha 12 activates Rho GTPase through tyrosine-phosphorylated leukemia-associated RhoGEF." Proc Natl Acad Sci U S A **100**(2): 733-8.
- Swiercz, J. M., R. Kuner, et al. (2002). "Plexin-B1 directly interacts with PDZ-RhoGEF/LARG to regulate RhoA and growth cone morphology." Neuron **35**(1): 51-63.
- Taya, S., N. Inagaki, et al. (2001). "Direct interaction of insulin-like growth factor-1 receptor with leukemia-associated RhoGEF." J Cell Biol **155**(5): 809-20.
- Tochio, H., Y. K. Mok, et al. (2000). "Formation of nNOS/PSD-95 PDZ dimer requires a preformed beta-finger structure from the nNOS PDZ domain." J Mol Biol **303**(3): 359-70.
- Togashi, H., K. Nagata, et al. (2000). "Functions of a rho-specific guanine nucleotide exchange factor in neurite retraction. Possible role of a

-
- proline-rich motif of KIAA0380 in localization." J Biol Chem **275**(38): 29570-8.
- Turner, L. J., S. Nicholls, et al. (2004). "The activity of the plexin-A1 receptor is regulated by Rac." J Biol Chem **279**(32): 33199-205.
- Ullmer, C., K. Schmuck, et al. (1998). "Cloning and characterization of MUPP1, a novel PDZ domain protein." FEBS Lett **424**(1-2): 63-8.
- van Golen, K. L., L. Bao, et al. (2002). "Reversion of RhoC GTPase-induced inflammatory breast cancer phenotype by treatment with a farnesyl transferase inhibitor." Mol Cancer Ther **1**(8): 575-83.
- Vasioukhin, V., C. Bauer, et al. (2000). "Directed actin polymerization is the driving force for epithelial cell-cell adhesion." Cell **100**(2): 209-19.
- Villalonga, P., R. M. Guasch, et al. (2004). "RhoE inhibits cell cycle progression and Ras-induced transformation." Mol Cell Biol **24**(18): 7829-40.
- Wallar, B. J. and A. S. Alberts (2003). "The formins: active scaffolds that remodel the cytoskeleton." Trends Cell Biol **13**(8): 435-46.
- Wang, Q., M. Liu, et al. (2004). "Thrombin and lysophosphatidic acid receptors utilize distinct rhoGEFs in prostate cancer cells." J Biol Chem **279**(28): 28831-4.
- Wells, C. D., S. Gutowski, et al. (2001). "Identification of potential mechanisms for regulation of p115 RhoGEF through analysis of endogenous and mutant forms of the exchange factor." J Biol Chem **276**(31): 28897-905.
- Wells, C. D., M. Y. Liu, et al. (2002). "Mechanisms for reversible regulation between G13 and Rho exchange factors." J Biol Chem **277**(2): 1174-81.
- Wen, Y., C. H. Eng, et al. (2004). "EB1 and APC bind to mDia to stabilize microtubules downstream of Rho and promote cell migration." Nat Cell Biol **6**(9): 820-30.
- Westwick, J. K., Q. T. Lambert, et al. (1997). "Rac regulation of transformation, gene expression, and actin organization by

- multiple, PAK-independent pathways." Mol Cell Biol **17**(3): 1324-35.
- Wiedemann, U., P. Boisguerin, et al. (2004). "Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides." J Mol Biol **343**(3): 703-18.
- Winter, C. G., B. Wang, et al. (2001). "Drosophila Rho-associated kinase (Drok) links Frizzled-mediated planar cell polarity signaling to the actin cytoskeleton." Cell **105**(1): 81-91.
- Wojciak-Stothard, B., S. Potempa, et al. (2001). "Rho and Rac but not Cdc42 regulate endothelial cell permeability." J Cell Sci **114**(Pt 7): 1343-55.
- Worthylake, R. A., S. Lemoine, et al. (2001). "RhoA is required for monocyte tail retraction during transendothelial migration." J Cell Biol **154**(1): 147-60.
- Xu, X. Z., A. Choudhury, et al. (1998). "Coordination of an array of signaling proteins through homo- and heteromeric interactions between PDZ domains and target proteins." J Cell Biol **142**(2): 545-55.
- Yamaguchi, Y., H. Katoh, et al. (2003). "N-terminal short sequences of alpha subunits of the G12 family determine selective coupling to receptors." J Biol Chem **278**(17): 14936-9.
- Yasuda, S., F. Ocegüera-Yanez, et al. (2004). "Cdc42 and mDia3 regulate microtubule attachment to kinetochores." Nature **428**(6984): 767-71.
- Yau, D. M., N. Yokoyama, et al. (2003). "Identification and molecular characterization of the G alpha12-Rho guanine nucleotide exchange factor pathway in *Caenorhabditis elegans*." Proc Natl Acad Sci U S A **100**(25): 14748-53.
- Zhou, K., Y. Wang, et al. (1998). "Guanine nucleotide exchange factors regulate specificity of downstream signaling from Rac and Cdc42." J Biol Chem **273**(27): 16782-6.

- Zimmermann, P., K. Meerschaert, et al. (2002). "PIP(2)-PDZ domain binding controls the association of syntenin with the plasma membrane." Mol Cell **9**(6): 1215-25.
- Zinovyeva, M., E. Sveshnikova, et al. (2004). "Molecular cloning, sequence and expression pattern analysis of the mouse orthologue of the leukemia-associated guanine nucleotide exchange factor." Gene **337**: 181-8.

2. Materials and methods

2.1. Polymerase Chain Reaction (PCR)

25pmol of each primer, 0.2mM dNTP mix, and 1mM MgCl₂, were used unless otherwise stated. Different standard PCR programmes were used for genomic DNA or plasmid DNA template. For genomic DNA, the following programme was used:

- | | |
|---|----------------|
| 1) 95 °C | 2 minutes |
| 2) 95 °C | 30 seconds |
| 3) {annealing temperature + 10 °C} – 2°C each cycle | 30 seconds |
| 4) 72 °C | extension time |
| 5) go back to step 2, repeat 5 times | |
| 6) 95 °C | 30 seconds |
| 7) annealing temperature | 30 seconds |
| 8) 72 °C | extension time |
| 9) go back to step 6, repeat 29 times | |
| 10) 72 °C | 10 minutes |
| 11) 10 °C | HOLD |

For plasmid DNA, the following programme was used:

- | | |
|---------------------------------------|----------------|
| 1) 95 °C | 2 minutes |
| 2) 95 °C | 30 seconds |
| 3) annealing temperature | 30 seconds |
| 4) 72 °C | extension time |
| 5) go back to step 2, repeat 29 times | |
| 6) 72 °C | 10 minutes |
| 7) 10 °C | HOLD |

The extension time was determined by the polymerase, following manufacturers instruction. Polymerases used included HOTStart Taq (Qiagen), Pfu Turbo (Stratgene), Expand long-template polymerase (Roche), Taq (Qiagen). The annealing temperature used was 1 °C lower

than the lowest melting temperature of the primers (for a list of primers used see **Appendix 2**). Generally, 150ng of DNA template were used per PCR. PCR products were checked by gel electrophoresis, purified using Qiaquick gel extraction kit (Qiagen) and sequenced where necessary (MWG Biotech).

2.2. Restriction digestion and ligation of DNA

Unless otherwise stated, 1µg DNA was digested using 5 units of restriction enzyme (New England Biolabs) at 37°C for 2 hours. Vector DNA was dephosphorylated by addition of 1 unit calf intestine phosphatase (Fermentas Life Sciences) to the digestion reaction and a further hour incubation at 37°C. Vectors used for cloning included: pGEMT Easy (Promega), pFASTBAC-Act5C (kindly donated by Buzz Baum), pBluescript (Stratagene), PCR Blunt (Invitrogen), pUASp (Rorth 1998). Digests were run on agarose gels and the DNA cleaned using Qiaquick gel extraction kit (Qiagen). Ligations were carried out using T4 DNA ligase (Invitrogen) according to manufacturer's instructions.

2.3. Transformation of bacteria

An aliquot of competent bacteria (Turbo cells, Invitrogen) was defrosted on ice. 50ng DNA was added and the mixture tapped gently. The mixture was incubated on ice for 5 minutes, heat-shocked at 42°C for 45 seconds, then placed immediately back on ice for 2 minutes. SOC medium (0.5% yeast extract, 2% tryptone, 10mM Na Cl, 2.5mM KCl, 10mM MgCl₂, 20mM MgSO₄, 20mM glucose) was added and the bacteria grown at 37°C for one hour. The bacteria were plated onto agar plates containing the relevant antibiotic and incubated at 37°C overnight.

Blue/white screening was performed where relevant using agar plates with 20mg/ml X-gal and 100mg/ml Isopropyl βD-thiogalactopyranoside (IPTG) smeared onto the surface.

2.4. Diagnostic PCR

PCRs of twelve bacterial colonies were performed using appropriate primers both sides of the multiple cloning site. PCR products were run on a 1% agarose gel. A small amount of each colony was transferred to a new agar plate before the tip was dipped in the PCR mix. The agar plate was incubated for ~8 hours to generate new colonies. Colonies giving the predicted PCR product size were selected for mini-prep.

2.5. Isolation of plasmid DNA from bacteria

Bacterial colonies were individually picked using a metal wire loop. These were used to inoculate LB broth (2% tryptone, 0.5% yeast extract, 10mM NaCl) cultures which were grown at 37 °C overnight. The Wizard *plus* SV Minipreps kit (Promega) was used to isolate the plasmid DNA on a small scale, and the Qiagen maxi prep kit on a large scale.

2.6. Generation of RNAi hybrid construct

The cDNA and genomic DNA pieces were generated by PCR. These were cloned through various intermediates and finally into the Not I and Xba I sites of pUASp (Rorth 1998).

2.7. Microinjection of the hybrid RNAi construct to generate transgenic flies

Transgenic flies were generated by P-element transformation (Rubin and Spradling 1982). One hour egg collections were performed using *yw* adult flies in an egg-laying cage. Embryos were dechorionated in 50% bleach for two minutes followed by three washes in dH₂O. Embryos were lined up on agar and transferred to double-sided sticky tape on a glass slide. Voltalef oil (Atochem) was used to cover the embryos and 1.5µg/µl DNA injected into the posterior end at either 3:1 or 5:1 ratio of UAS-MCC RNAi: Turbo transposase.

2.8. Selection and balancing of RNAi transgenic flies

Surviving adult flies were crossed to yw flies. Second generation flies that had mw⁺ eyes (i.e. pale yellow through to red) were selected and individually crossed back to yw. Males were also crossed to X, second and third chromosome balancer virgins once they had fertilised the females of the previous cross. All subsequent generations were crossed back to balancers to determine on which chromosome the insertion(s) lay. Lines were maintained as balanced stocks or homozygotes if viable as such.

2.9. Transgenic RNAi and UAS-MCC crosses to GAL4 lines

pUASp-MCC RNAi or pUAST-MCC lines were crossed to various GAL4 drivers at 25°C (table 2.1). Flies were scored soon after hatching. All RNAi, over-expression, and GAL4 lines were homozygous viable, therefore all progeny carried both insertions and were scored phenotypically.

2.10. RNA extraction for RT-PCR

20 adult flies were rapidly homogenised with a pestle in 700µl (1 volume) guanidium hydrochloride solution in a 1.5ml Eppendorf tube. An equal volume of 50% phenol:50% chloroform was added and mixed well by vortexing. The sample was centrifuged at 12,000 x g for 5 minutes at 4°C. The upper, aqueous phase was transferred to a new tube. The RNA was precipitated by addition of 0.02 volumes 1M acetic acid and 0.5 volumes 100% ethanol. Following gentle mixing the sample was incubated at -20°C for at least 3 hours. The RNA was pelleted by centrifugation at 14,000 x g for 10 minutes at 4°C. The pellet was redissolved in 0.5 volumes guanidium hydrochloride solution and re-precipitated as above. To the RNA pellet 1 volume of 100% ethanol was added to wash.

Driver name	Expression pattern	Reference
{tubP-GAL4}LL7	Ubiquitous	(Lee and Luo 1999)
{GawB}elav ^{C155}	All nervous system tissues from stage 12	(Lin and Goodman 1994)
{GAL4-wg.M}MA1	Wingless (striped) expression pattern in embryos	(Hays, Gibori et al. 1997)
{GAL4-ey.H}	Eye	(Hazelett, Bourouis et al. 1998)
{sevEP-GAL4.B}	Eye	(Therrien, Wong et al. 1999)
GMR-GAL4	Eye	(Moses and Rubin 1991)
Da-GAL4	Ubiquitous (strong)	(Georgias, Wasser et al. 1997)

Table 2.1. GAL4 drivers.

Following centrifugation at 14,000 x g for 5 minutes at 4°C the pellet was air dried for approximately 5 minutes and resuspended in DEPC-treated ddH₂O. The sample was stored at -20°C for short-term storage and -80°C for long term storage.

Guanidium hydrochloride solution:

5mM DTT
7.5M guanidium hydrochloride
25mM sodium acetate, pH 7.0
0.5% N-lauryl sarcosinate

DEPC-treated ddH₂O:

0.1% diethyl pyrocarbonate in ddH₂O
left overnight and autoclaved

2.11. Reverse transcription and PCR

Reverse transcription was carried out using the Single Strand RT-PCR kit (Invitrogen) following manufacturer's instructions including DNase I treatment prior to reverse transcription. Reactions were primed using oligo dT.

2.12. Probe generation for whole mount *in situ* hybridisation

3µg probe DNA in the pBluescript vector with T7 and SP6 RNA polymerase priming sites was digested with appropriate restriction enzymes to generate sense (control) and anti-sense probes by *in vitro* transcription using the DIG RNA labelling kit (Roche).

2.13. *In situ* hybridisation

An overnight collection of embryos was dechorionated in 50% bleach and washed thoroughly with ddH₂O before being transferred to a polypropylene tube containing equal volumes of fixative and heptane, and shaken at room temperature for 30 minutes. The lower phase (fixative) was removed and an equal volume of methanol added. Devitellinsation was achieved by shaking vigorously in heptane/methanol for 30 seconds. The devitellinised embryos sank to the bottom of the methanol phase and were recovered to a fresh tube.

The embryos were washed three times with methanol and rehydrated gradually into 4% paraformaldehyde in PBS followed by re-fixing in 4% paraformaldehyde for 20 minutes at room temperature. They were then washed in PTw three times for 5 minutes each at room temperature and incubated in an equal mix of PTw and hybridisation solution for 10 minutes, followed by a further 10 minute incubation in pure hybridisation solution. Pre-hybridisation was carried out at 55°C for one hour. The probe was boiled for 10 minutes at 100°C then placed on ice for 1 minute. Fresh hybridisation solution containing probe at 500pg/µl was added and the embryos were incubated in probe overnight at 55 °C.

Washes at 55 °C in hybridisation solution for 20 minutes were followed by subsequent washes in gradually increasing concentration of PTw. The embryos were washed in pure PTw twice at room temperature followed by two washes in PBT. They were then incubated for one hour in 1: 2500 anti-digoxigenin antibody (Roche) at room temperature followed by four washes with PBT and three with detection buffer at room temperature. The embryos were transferred to a glass dish to facilitate visualisation

and incubated in detection buffer containing 3.75 µg/ml Nitro blue tetrazolium chloride and 1.88 µg/ml 5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt. Once the colour reaction was complete, the embryos were washed three times with PBS then placed in an Eppendorf tube in 70% glycerol to clear overnight.

Fixative

4% paraformaldehyde

60mM Na₂HPO₄

40mM NaH₂PO₄

Hybridisation Buffer

50% deionised formamide

5 x SSC

100µg/ml denatured phenol-chloroform extracted salmon sperm DNA

100µg/ml heparin

0.1% Tween 20

PTw

1 x PBS

0.1% Tween 20

50mM MgCl₂

0.1% Tween 20

Detection Buffer

100mM NaCl

100mM Tris-Cl pH 9.5

PBT

1 x PBS

0.1% Bovine serum albumin

0.2% Triton X-100

2.14. Plasmid rescue of P(lacW) insertions

Genomic DNA was extracted from twenty flies and digested with Xba (5' end of P-element rescued) or ECoRI (3' end of P-element rescued). The restriction enzyme was heat-inactivated at 70°C for 10 minutes and the volume was made up to 200µl with ddH₂O. The restriction fragments were ligated overnight at 4 °C using T4 DNA ligase. The DNA was precipitated by addition of 40ng tRNA, 2 volumes 100% ethanol and 0.1

volumes 3M sodium acetate. The precipitation reaction was incubated at -20°C for 30 minutes and centrifuged at 13,000 x g for 30 minutes at 4°C. The supernatant was removed and the pellet washed with 70% ethanol followed by air drying and resuspension in ddH₂O.

The DNA was transformed into bacteria, colonies were grown up as mini preps and sequencing reactions carried out. The resulting sequence was lined up against the *Drosophila* genome using BLAST in order to determine the site of P-element insertion.

2.15. P-element hop genetics

Drosophila with a P[lacW^{mw}] P-element were obtained from the Szeged Stock Centre (stock no. 070116). After three generations in quarantine these were crossed with a source of transposase (Robertson, Preston et al. 1988) in order to generate flies with P-element insertions at new positions.

Cross 1

♂		♀
yw ; Ki p ^p Δ2-3	x	yw ; P[lacW ^{mw}]
		yw TM3 Sb, Ser

Cross 2

Mosaic-eyed male flies were selected and these were crossed to a third chromosome balancer stock:

♂		♀
yw ; P[lacW ^{mw}]	x	w ; TM2 Ubx
Ki p ^p Δ2-3		w TM6C Sb

Cross 3

Males with *mw+* were selected and crossed again to the third chromosome balancer stock in order to balance the new P-element insertions:

♂		♀
w ; <u>P[lacW^{mw}]</u>	x	<u>w</u> ; <u>TM2 Ubx</u>
TM2 Ubx or TM6CSb		w TM6C Sb

Flies with insertions on chromosomes other than the third were de-selected after cross 3 due to the presence of both balancer chromosomes and *mw+*. Flies were maintained as balanced stocks during PCR screening.

2.16. Genetic interaction studies

MCC^{1,2,3} represent different MCC alleles e.g. the P1 insertion, P4 insertion etc.. RG2^{1,2,3} represent different DRhoGEF2 alleles e.g. DRhoGEF2^{4.1}, DRhoGEF2^{Px6} etc.....(see **table 5.3**).

1) Heterozygous DRhoGEF2 with heterozygous DMCC.

Cross 1:

♂		♀
<u>yw</u>	x	<u>w</u> ; + ; <u>MCC¹</u>
		w + TM6C Sb

Cross 2:

♂		♀
<u>yw</u> ; <u>MCC¹</u>	x	<u>w</u> ; <u>RG2¹</u> ; <u>±</u>
+		w CyO +

2) Homozygous DRhoGEF2 with heterozygous DMCC.

Cross 1:

♂		♀
<u>w</u> ; <u>RG2</u> ¹ ; ±	x	<u>w</u> ; <u>If</u> ; <u>MCC</u> ¹
CyO +		w CyO TM6C Sb

Cross 2:

♂		♀
<u>w</u> ; <u>RG2</u> ² ; ±	x	<u>w</u> ; <u>RG2</u> ¹ ; <u>MCC</u> ¹
CyO +		w CyO +

3) Homozygous DRhoGEF2 with homozygous DMCC.

Cross 1:

♂		♀
<u>w</u> ; <u>RG2</u> ¹ ; ±	x	<u>w</u> ; <u>If</u> ; <u>MCC</u> ¹
CyO +		w CyO TM6C Sb

Cross 2:

♂		♀
<u>w</u> ; <u>RG2</u> ² ; <u>MCC</u> ²	x	<u>w</u> ; <u>RG2</u> ¹ ; <u>MCC</u> ¹
CyO +		w CyO +

2.17. Genomic DNA isolation for PCR screening

Single flies from each line were pooled in groups of twenty for genomic DNA isolation. The flies were homogenised in 250µl homogenisation buffer and an equal volume of lysis buffer was added and mixed by inversion. RNase A was added to 25µg/ml, mixed by inversion, and incubated at 37°C for 15 minutes. Proteins were precipitated by addition of proteinase K to 400µg/ml, which was mixed by inversion, and incubated at 37°C for 30 minutes. 1 volume of 50% phenol/50% chloroform was added and mixed vigorously by shaking. The sample

was centrifuged at 13,000 x g for 10 minutes at 4°C and the upper aqueous phase recovered to a fresh tube. The phenol/chloroform extraction was repeated. 0.7 volumes of pure chloroform was added, mixed by inversion and the sample was centrifuged at 13,000 x g for 10 minutes at 4°C. The upper aqueous phase was recovered to a fresh tube.

2.5 volumes of 100% ethanol was added, mixed by inversion, and the sample was centrifuged at 13,000 x g for 5 minutes at 4°C. The pellet was washed with 70% ethanol. The pellet was air-dried for 5-10 minutes, resuspended in ddH₂O and stored at 4°C.

Homogenisation buffer:

10mM Tris-Cl pH 7.5
60mM Na Cl
10mM EDTA
0.15mM spermine
0.15mM spermidine
5% sucrose

Lysis buffer:

300mM Tris-Cl pH 9.0
100mM EDTA
0.626% SDS
5% sucrose

2.18. Digoxigenin-DNA labelling of probes for Southern blotting

Probes were labelled by random primer labelling following manufacturer's instructions using the DIG DNA Labelling and Detection Kit (Roche).

1.5µg probe DNA was used in the labelling reaction. The labelled DNA was precipitated using 0.5M LiCl and 3.75 volumes ice-cold 100% ethanol. The sample was incubated at -20°C for 30 minutes then centrifuged at 13000 x g for 15 minutes at 4°C. The pellet was washed with ice cold 70% ethanol, dried at 45°C for 10 minutes and redissolved in TE.

2.19. Dot blot test

A dot blot test was performed to determine the concentration of the labelled probe. A series dilution (from undiluted to 1 in 100,000) of the labelled probe and the control labelled probe was performed in ddH₂O. 1 µl of each dilution was dot-blotted onto a Hybond-N nylon membrane (Amersham Life Sciences). The control labelling reaction and the control pre-labelled DNA were also blotted. The DNA was fixed to the membrane by UV cross-linking using the UV Stratalinker 2400 (Stratagene) at 120 Joules/cm².

The membrane was washed in maleic acid buffer with agitation for 15 minutes at room temperature followed by incubation in 1 x blocking solution for 30 minutes with agitation. It was then incubated in 1:5,000 anti-DIG-AP antibody (Roche) in 1 x blocking solution for 30 minutes followed by two 15 minute washes with washing buffer then a 5 minute incubation in detection buffer. Spots were detected by incubation in detection buffer containing 3.75 µg/ml Nitro blue tetrazolium chloride and 1.88 µg/ml 5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt in the dark. The reaction was stopped by addition of excess ddH₂O and the membrane was air-dried.

2.20. DNA digestion and denaturation for southern blot

1.5 µg genomic DNA was digested for 2 hours at 37°C with a suitable 6-base cutter restriction enzyme. The DNA was run on a 1% agarose gel containing ethidium bromide at 40V for approximately 8 hours. The DNA was visualised under ultraviolet light to check digestion was complete. The gel was incubated in denaturing solution for 30 minutes followed by neutralising solution for 30 minutes at room temperature with agitation.

Denaturing Solution:

1.5M NaCl
1M Tris-Cl pH 8.0

Neutralising Solution:

1.5M Na Cl
0.5M NaOH

2.21. Transfer of DNA to membrane and probe hybridisation

The DNA was transferred overnight from gel to Hybond –N nylon membrane (Amersham Life Sciences) as in (Sambrook J 1989) and cross-linked to the membrane as for dot blot. The membrane was rinsed in 3 x SSC and placed in a glass hybridisation bottle. The membrane was pre-hybridised at 42°C for 2-4 hours in hybridisation solution with rotation. Hybridisation solution was removed and replaced with fresh solution containing 100ng probe. Hybridisation was carried out at 42°C overnight.

2.22. Stringency washes and detection

The probe solution was drained off and stored for future use at -20°C. The membrane was washed flat with agitation with excess wash solution 1 at room temperature. The membrane was washed flat with agitation in excess wash solution 2 at 68°C. All subsequent steps were performed at room temperature unless otherwise stated. The membrane was incubated in 1 x blocking solution for 30 minutes with agitation followed by 30 minute incubation in 1:5,000 anti-DIG-AP antibody in 1 x blocking solution. It was then washed twice with washing buffer followed by a 5 minute incubation in detection buffer. 125µM Disodium 3-(4-methoxy-spiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo [3.3.1.1^{3,7}] decan}-4-yl)phenyl phosphate (CSPD) in detection buffer was added and incubated at 37°C for 10 minutes. The membrane was exposed to photo-sensitive film for ~30 minutes and developed.

20X SSC:

3M NaCl

0.3M Na citrate

pH 7.5 (pH adjusted using solid NaOH)

Wash solution 1:

2 X SSC

0.1% SSC

Maleic Acid Buffer:

0.1M maleic acid

0.15M NaCl

10 X Blocking Solution:

10% blocking reagent (Roche)

in maleic acid buffer

This is diluted 1 in 10 with maleic acid buffer to produce working solution

Wash solution 2:

0.1X SSC

0.1% SDS

Washing Buffer:

Maleic acid buffer

0.3% Tween 20

Detection Buffer:

0.1M Tris-Cl pH 9.5

0.1M NaCl

Hybridisation Solution:

50% formamide

1x Denhardt's (0.02% Ficoll (type 400), 0.02% polyvinylpyrrolidone, and 0.02% bovine serum albumin)

1% SDS

5 X SSC

5% blocking reagent

0.1M sodium phosphate buffer

100ug/ml denatured salmon sperm DNA

2.23. Storing, stripping and re-probing the membrane

The membrane was stored at -20°C in sealed plastic casing. To strip the membrane of probe it was briefly rinsed in ddH₂O and incubated twice in strip buffer (0.2M Na OH, 0.1% SDS) for 15 mins at 37°C. The membrane was rinsed thoroughly in excess 2 x SSC, incubated in hybridisation solution as before, and hybridisation with a second probe carried out.

2.24. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE)

A 7.5% separating gel and stacking gel were prepared as described (Sambrook J 1989). The gel was loaded with protein samples and run at 95V until the blue marker dye had migrated to the bottom of the gel.

2.25. Transfer and Immunoblotting

Protein samples were transferred to nitrocellulose membrane as described (Sambrook J 1989). Following transfer the membrane was washed twice with TBS Tween for 5 minutes followed by blocking solution for 30 minutes. Primary antibody in blocking solution was added to a final concentration of 1:1000 (anti-myc^{9E10}) or 1:10,000 (anti-T7 tag) and the blot incubated overnight at 4°C with gentle shaking. The blot was washed 3 times for 10 minutes each at room temperature with TBS Tween. Goat anti mouse-HRP secondary antibody (DAKO) was added in blocking solution at 1:2000 and incubated at room temperature for 1 hour. The blot was washed 3 times for 10 minutes each at room temperature with TBS Tween. The blot was developed and visualised following manufacturer's instructions using ECL blotting detection reagents (Amersham Biosciences).

TBS Tween:

10mM Tris-Cl pH 8.0
150mM Na Cl
0.1% Tween 20

Blocking solution:

5% milk powder (Tesco) in
TBS Tween

2.26. Stripping and reprobing membrane

The membrane was washed briefly in TBS Tween, strip buffer was added and incubated with shaking for 30-60 mins at 65°C. The membrane was washed several times in TBS Tween before proceeding to block and antibody incubation as above.

Strip buffer:

2% SDS

62.5mM Tris-Cl pH 6.8

100mM β -mercaptoethanol

2.27. Tagged DRhoGEF2 and MCC construct preparation

Full-length DRhoGEF2 was previously cloned into pUASp with a T7 tag at the N-terminus (K. Barrett). pFASTBAC-act5C was generated by replacement of the polyhedron promoter of pFastBac 1 (Invitrogen) with the actin 5C promoter, and was a kind gift from Buzz Baum. pFastBac-act5C-myc was generated by annealing two oligos (MWG) to create the myc tag with an upstream Kozak sequence and sticky restriction sites on each end (**appendix 2**).

200pmol of each primer were annealed in restriction enzyme buffer (New England Biolabs) in a PCR machine using the following program: 90°C for 5 minutes, 70°C for 10 minutes, -2°C each minute until 4°C. The newly double-stranded DNA was phosphorylated using 10 units polynucleotide kinase (Fermentas) in a forward reaction following manufacturer's instruction. The fragment was ligated into restriction-digested pFastBac-act5C. MCC constructs were generated by RT PCR followed by cloning into pGEMT Easy vector (Promega). PCR using primers with appropriate restriction sites was used to transfer MCC constructs from pGEMT easy into pFastBac-act5C-myc. Mutant MCC constructs in pCMV vector were provided by Kathy Barrett. These were sub-cloned into pFASTBAC-act5C-myc.

2.28. Transfection and lysis of *Drosophila* cells

S2 cells were grown at high density in Schneider's medium with 10% fetal calf serum and transfected according to manufacturer's instructions using Fugene transfection reagent (Roche). At 2 days post-transfection cells

were spun down at 1000 rpm and lysed by pipetting up and down in standard lysis buffer

Lysis Buffer:

20mM Tris-CL pH 7.5

1% NP40

150mM NaCl

1mM PMSF

10µg/ml leupeptin

10µg/ml aprotinin

2.29. Full-length protein co-immunoprecipitation

A 10cm dish of cells was transfected as above using 6µg DRhoGEF2, 6µg actin-GAL4, 6µg MCC DNA. Cells were lysed in 400µl lysis buffer. Where phosphatase inhibitors were added the lysis buffer included 1mM sodium orthovanadate, 20mM sodium fluoride and 12.5mM sodium pyruvate. The sample was centrifuged at 10,000 rpm for 5 minutes at 4°C. 20 µl post-nuclear supernatant samples were transferred to a new tube and 20 µl 2 x sample buffer added and mixed. These were stored at -20°C. The remaining lysate was transferred to a new tube and EDTA and GTPγS (Upstate) were added where appropriate to 1mM and 100µM concentrations respectively, and the sample incubated at 30°C for 5 minutes with agitation. The reaction was stopped by the addition of magnesium chloride to a final concentration of 60mM.

2 x sample buffer:

100mM Tris-Cl pH 6.8

4% SDS

0.2% bromophenol blue

20% glycerol

5mM DTT

Bead preparation - 40 μ l protein G beads (Autogen Bioclear) per sample were spun down at 1,000 x g and washed three times with cold PBS. 20 μ l per sample were used for clearing. Primary antibody was added to the remaining 20 μ l per sample (mouse anti-myc 9E10 (Santa Cruz Biotechnology) was used at 1:1000, mouse anti-T7 tag (Novagen) was used at 1:5000) and the beads were incubated with end-over-end mixing at 4 °C for 30 minutes.

The lysate was cleared by end-over-end incubation with 20 μ l pre-washed protein G beads at 4 °C for 30 minutes. Following clearing the beads were spun down and the lysate recovered to a fresh tube. 20 μ l antibody-conjugated beads were added to the lysate and the sample was end-over-end mixed for at least 1 hour at 4 °C.

The beads were washed 3 times with 1ml lysis buffer. Each wash involved addition of wash buffer, brief end-over-end mixing followed by centrifugation at 14,000 rpm for 1 minute at 4 °C. Beads were resuspended in 20 μ l 2 x sample buffer, boiled at 100 °C for 5 minutes, spun down and the supernatant loaded onto a 7.5% SDS PAGE gel (or stored at -20 °C). The blot was immunoblotted with the appropriate antibody to test for immunoprecipitation and co-immunoprecipitation.

2.30. Recombinant PDZ domain production

The PDZ domain of DRhoGEF2 was cloned into the PET28c vector (Novagen). BL21 cells (Novagen) were transformed with the construct and a single colony was used to inoculate 50ml LB broth. The culture was incubated with shaking at 37°C until the OD₆₀₀ reached 0.6 when IPTG was added to a final concentration of 1mM. The culture was grown for a further 3 hours and the bacteria were then spun down at 10,000 x g for 10 minutes. The pellet was resuspended in 4ml ice-cold 20mM Tris-Cl pH 7.5. rLysozyme (Novagen) was added to a final concentration of 50KU/gram bacterial pellet and the cell suspension was sonicated on ice using a microtip with amplitude 60 for 10 second bursts until the

viscosity was significantly reduced (approx 10 x 10 second bursts). The sample was centrifuged at 14,000 x g for 10 minutes to separate the insoluble phase which was discarded. The soluble supernatant was passed through a resin column and purified according to manufacturer's instructions using the His Bind Purification Kit (Novagen). Samples were run on SDS-PAGE gels and stained with Coomassie blue as described (Sambrook J 1989).

2.31. PDZ domain co-immunoprecipitation

60mm plates of S2 cells were transfected as above with individual MCC constructs. Cells were lysed in 200µl standard lysis buffer and 20µl was kept as whole cell lysate. Protein A beads were prepared as above and the lysate was cleared using T7-conjugated beads as above. 500ng recombinant PDZ domain was incubated with T7-conjugated beads for 1hour at 4°C prior to addition to the cleared cell lysate, followed by overnight incubation at 4°C. The samples were washed as above, run on an SDS-PAGE gel and immunoblotted.

2.32. Chapter 2 References

- Georgias, C., M. Wasser, et al. (1997). "A basic-helix-loop-helix protein expressed in precursors of *Drosophila* longitudinal visceral muscles." Mech Dev **69**(1-2): 115-24.
- Hays, R., G. B. Gibori, et al. (1997). "Wingless signaling generates pattern through two distinct mechanisms." Development **124**(19): 3727-36.
- Hazelett, D. J., M. Bourouis, et al. (1998). "decapentaplegic and wingless are regulated by eyes absent and eyegone and interact to direct the pattern of retinal differentiation in the eye disc." Development **125**(18): 3741-51.
- Lee, T. and L. Luo (1999). "Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis." Neuron **22**(3): 451-61.

- Lin, D. M. and C. S. Goodman (1994). "Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance." Neuron **13**(3): 507-23.
- Moses, K. and G. M. Rubin (1991). "Glass encodes a site-specific DNA-binding protein that is regulated in response to positional signals in the developing *Drosophila* eye." Genes Dev **5**(4): 583-93.
- Robertson, H. M., C. R. Preston, et al. (1988). "A stable genomic source of P element transposase in *Drosophila melanogaster*." Genetics **118**(3): 461-70.
- Rorth, P. (1998). "Gal4 in the *Drosophila* female germline." Mech Dev **78**(1-2): 113-8.
- Rubin, G. M. and A. C. Spradling (1982). "Genetic transformation of *Drosophila* with transposable element vectors." Science **218**(4570): 348-53.
- Sambrook J, F. E., Maniatis T (1989). "Molecular Cloning: A Laboratory Manual (2nd ed.)." Cold Spring Harbor Laboratory Press, New York.
- Therrien, M., A. M. Wong, et al. (1999). "Functional analysis of CNK in RAS signaling." Proc Natl Acad Sci U S A **96**(23): 13259-63.

3. Results – DMCC expression analysis

3.1. Introduction

When considering the function of MCC in *Drosophila* development a good place to begin analysis is to determine where and when the gene is expressed and determine the nature of the protein products resulting from gene expression. Sequence analysis of the mRNA transcribed from a gene allows prediction of the primary sequence of the resulting protein, and this protein sequence can be used to make predictions about function. Analysis of the conservation of a protein across species can also provide information. If a protein is expressed across many species, and particularly if it has functional domains that are well-conserved, this indicates that the protein is very likely to play an important role in survival of the species.

The expression pattern of a particular mRNA gives clues as to the expression pattern of the protein (the two do not necessarily coincide exactly), and can provide information as to the likely function of the protein. If an mRNA is only expressed in one particular tissue, for example, it is highly likely that the corresponding protein will be functioning in that tissue, and some inference of function may be possible. In this chapter the structure and expression of the *Drosophila* MCC gene is explored.

3.2. The protein encoded by the predicted gene CG6156 interacts with DRhoGEF2 PDZ domain

A yeast-2-hybrid screen using the DRhoGEF2 PDZ domain identified the protein product of a mRNA produced by the CG6156 locus of the *Drosophila* genome as a potential PDZ target (K. Barrett, unpublished). This gene is predicted to encode two mRNA isoforms corresponding to cDNAs GH12452 and GH21874 (subsequently referred to as cDNAa and cDNAb respectively), which were obtained from a Berkeley *Drosophila* Genome Project (BDGP) adult head cDNA library.

The mRNAs corresponding to cDNAa and cDNAb are comprised of 7 exons and 6 exons respectively, the difference being that the first intron is not removed during exon splicing of mRNA_b (**fig. 3.1**). This intron contains a STOP codon in frame with the mRNA_a transcription start site and it is therefore predicted that the protein encoded by mRNA_b will begin at the next downstream methionine which is just into the second exon. The two mRNAs also have slightly different sites of 3' cleavage and poly-adenylation but this is not expected to lead to a difference in protein sequence. mRNA_a and mRNA_b are predicted to encode proteins of 630 and 565 amino acids respectively with identical protein sequence except the 65 amino acid N-terminus of mRNA_a that is absent in mRNA_b (**appendix 1**). It is possible that a 19 amino acid polypeptide is also produced from mRNA_b as a result of translation up to the STOP codon in the first intron.

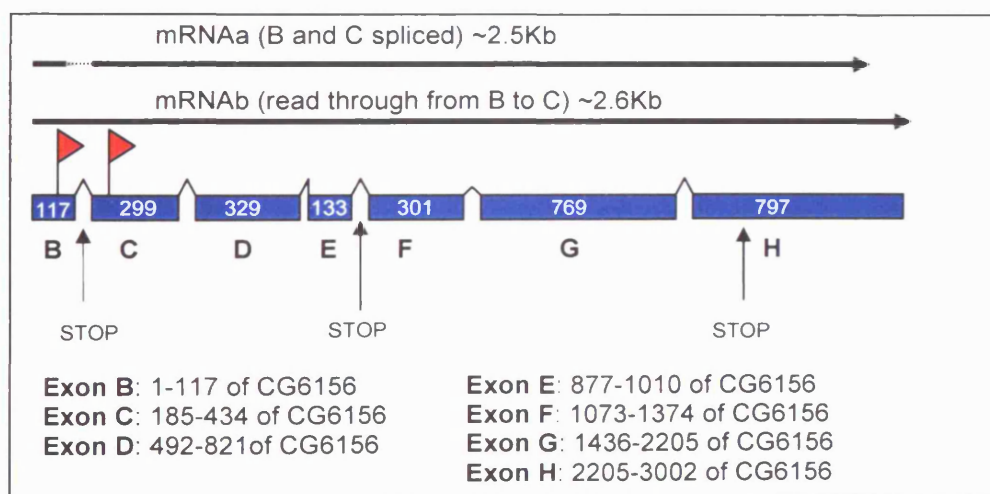


Figure 3.1. CG6156 mRNAs. mRNA_a and mRNA_b encode proteins with different N-termini. Blue boxes denote exons, red flags denote methionine (ATG) residues and STOP arrows denote stop codons. Numbers in exon boxes indicate size in bp. Numbers in table refer to bp of CG6156 published sequence (**appendix 1**). Not to scale.

3.3. The predicted protein encoded by CG6156 has sequence homology to human MCC

cDNAa and cDNAb do not have significant sequence homology to DNA sequences from either *Drosophila* or other species as determined using Basic Local Alignment Sequence Tool (BLAST) (Altschul, Madden et al. 1997). However, the predicted protein encoded by mRNAa (referred to as DMCCa from here onwards) has significant primary sequence homology to the human MCC1 protein in two regions (fig. 3.2). An arginine residue frequently mutated in human patients with colon cancer (Kinzler, Nilbert et al. 1991) is conserved in the fly protein.

A BLAST search using DMCCa identifies MCC2 as well as MCC1 (table 3.1). The reciprocal search using MCC1 picks up DMCC, but a search using MCC2 picks up nothing in *Drosophila*. This indicates that there is only likely to be one MCC gene in *Drosophila*, and that it has greater sequence similarity to MCC1 than MCC2.



Figure 3.2. Regions of sequence identity between human MCC1 and *Drosophila melanogaster* MCCa protein. Yellow highlights identical residues, red highlights Arg 506 of the human protein which is mutated in some colon cancer patients. Numbers refer to published amino acid sequences (appendix 1). Adapted from BLAST search result.

3.4. MCC has other species orthologues

DMCCa has predicted orthologues in other *Drosophila* species and in mammals including the mouse and orang-utan (**table 3.1**). A BLAST search does not identify any predicted orthologues in the nematode worm or yeast.

All of the species in **Table 3.1** (except *Drosophila pseudoobscura*, the crab-eating monkey, orang-utan, and non-animal species) have DRhoGEF2 protein orthologues identified in a BLAST search. In addition, DRhoGEF2 has orthologues in species for which there are not predicted MCC orthologues including the mosquito (*Anopheles gambiae*), the nematode worms (*Caenorhabditis elegans* and *Caenorhabditis briggsae*), the frog (*Xenopus laevis*) and the pufferfish (*Tetraodon nigroviridis*).

At least some of those animals that do not have identifiable orthologues of either DMCC or DRhoGEF2 have not had their genomes completely sequenced to date, and there could, therefore, be orthologues that are unidentifiable by BLAST at present.

Figure 3.3 illustrates the degree of homology between DMCC and its orthologues. The region of DMCC from amino acid residues ~280-500 has the highest degree of homology across species. This region includes sections of two predicted coiled coils and the conserved arginine residue (see **fig. 3.5**).

3.5. MCC has a conserved PDZ target motif at its C-terminus

Sequence analysis reveals that DMCCa, *Drosophila* mRNAb predicted protein (DMCCb) and human MCC (HsMCC) all have a consensus (threonine/serine-X-leucine) PDZ target motif at their C-terminus (where X represents any amino acid) (**fig. 3.4**).

	Protein name	Identity	Similarity	C terminus
Fruit fly (<i>Drosophila pseudoobscura</i>)	GA19398 (720aa)	77%	85%	PETTL
Honey bee (<i>Apis mellifera</i>)	Similar to MCC (1254aa)	28%	46%	KRANG
Human (<i>Homo sapiens</i>)	MCC (829aa)	27%	47%	NETSL
Orang-utan (<i>Pongo pygmaeus</i>)	Similar to MCC (766aa)	27%	47%	NETSL
House mouse (<i>Mus musculus</i>)	Similar to MCC (1303aa)	26%	46%	EEISN
Crab-eating monkey (<i>Macaque fascicularis</i>)	Similar to MCC (829aa)	27%	46%	NETSL
Red jungle fowl (<i>Gallus gallus</i>)	Similar to MCC (1454aa)	35%	56%	NETSL
Brown rat (<i>Rattus norvegicus</i>)	Similar to MCC (1133aa)	25%	41%	NETSL
Red jungle fowl (<i>Gallus gallus</i>)	Similar to AIEBP (MCC2) (685aa)	22%	44%	PCMRL
Cyanobacteria (<i>Nostoc sp.</i> PCC 7120)	Chromosome segregation protein (1208aa)	25%	48%	SNTSA
Head blight fungus (<i>Fusarium graminearum</i>)	FG00382.1 (1459aa)	22%	41%	RWLGG
Human (<i>Homo sapiens</i>)	AIEBP (MCC2) (703aa)	26%	44%	GDTFL
Chimpanzee (<i>Pan troglodytes</i>)	Similar to MCC (688aa)	27%	48%	NETSL
Rice blast fungus (<i>Magnaporthe grisea</i>)	MG00594.4 (2056aa)	24%	46%	NNATE

Table 3.1. DMCC potential orthologues. (previous page). A protein BLAST search using DMCCa predicted protein was performed. Results are ranked with lowest BLAST e-value first. Numbers in parentheses following protein name denote protein length in amino acids (aa).

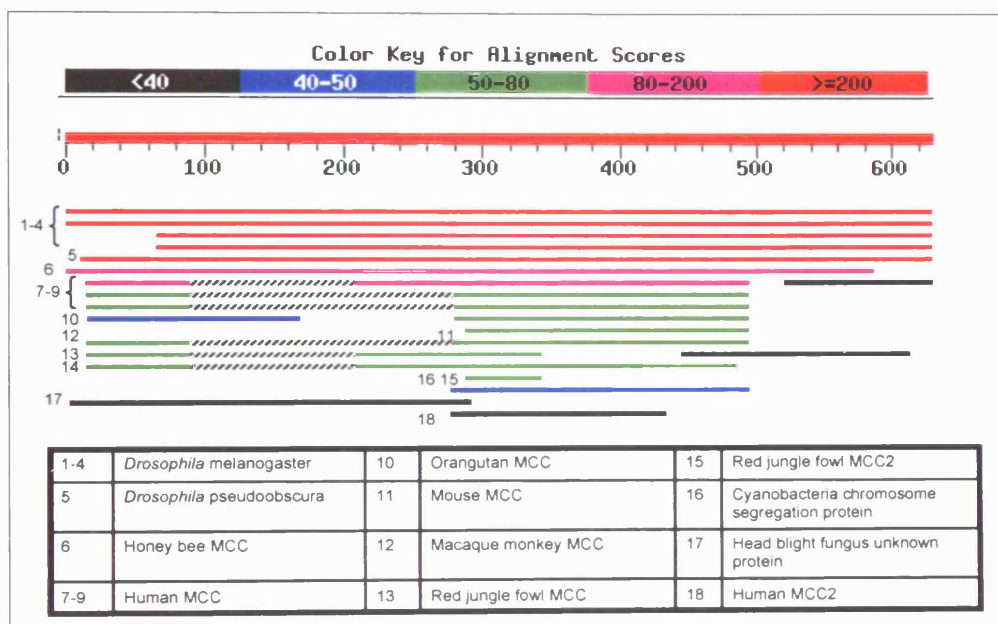


Figure 3.3. MCC similarity across species. Results of BLAST using DMCCa protein. Colours denote degree of homology between DMCC and species orthologues. Scores are arbitrary units.

This PDZ binding motif is conserved in the “similar to MCC” proteins of the crab-eating monkey, the orang-utan, the chimpanzee, the brown rat and the red jungle fowl (table 3.1). Neither the house mouse nor the honey bee “similar to MCC” proteins are predicted to contain a PDZ target motif at their C-termini. *Drosophila pseudoobscura* MCC contains exactly the same motif (threonine-threonine-leucine) at its C-terminus as DMCCa.

All the species which have DRhoGEF2 predicted orthologues and DMCC predicted orthologues including the human, the honey bee, the red jungle fowl, the house mouse, the brown rat and the chimpanzee have predicted PDZ domains in a DRhoGEF2 orthologue (as determined by Simple

Modular Architecture Research Tool (SMART) analysis (Letunic, Copley et al. 2004)).

-9	-8	-7	-6	-5	-4	-3	-2	-1	0	
G	H	N	H	V	P	E	T	T	L	Fly
S	R	P	H	T	N	E	T	S	L	Human

Figure 3.4. Fly and human MCC both have a classic C-terminal PDZ target sequence: T-X-L-COOH. Numbers denote amino acid position from the C-terminus which is position 0.

As noted above, of these the honey bee and the house mouse “similar to MCC” proteins do not carry a predicted PDZ target motif at their C-terminus. Therefore, at the current time, the species with an identifiable PDZ-containing-Rho GEF and a C-terminal target-containing-MCC are *Drosophila melanogaster*, red jungle fowl, brown rat, chimpanzee and human.

In addition, the nematode worm, frog and pufferfish have predicted PDZ domains in their DRhoGEF2 orthologues.

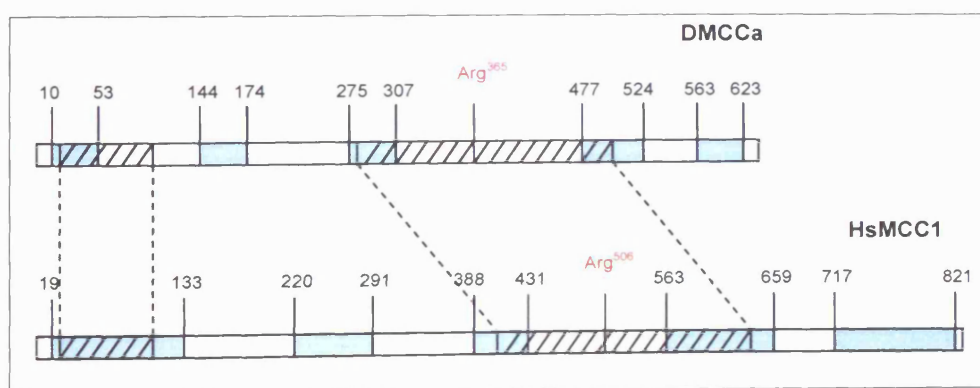


Figure 3.5. MCC has a repeat coiled coil structure. Blue regions indicate coiled coils. Dashed regions indicate significant sequence identity (fig. 3.2). Numbers denote amino acid sequence (appendix 1). Not to scale.

3.6. MCC proteins have a predicted repeat coiled coil structure

SMART predicts HsMCC and DMCCa to each contain five coiled coils (fig. 3.5). The two regions of significant sequence identity cover parts of regions predicted to form coiled coils. Arg⁵⁰⁶ does not fall within a predicted coiled coil region.

3.7. RT PCR analysis reveals that there are two additional *Drosophila* MCC isoforms

Previous work had indicated that there may exist, in addition to mRNAa and mRNAb, further mRNAs encoded by the CG6156 locus (V. Finnerty, personal communication). To address this possibility, reverse transcriptase PCRs were performed to amplify two novel cDNAs (figs. 3.6 and 3.7). The primers were designed to cover an exon from an upstream predicted gene that may be upstream of the CG6156 exons in a novel mRNA splice form, and a short sequence immediately following exon E that may be included in a separate splice form. For primer sequences see appendix 2. The primers used in the RT PCRs were:

cDNAa:	MCC RT PCR_2 & MCC RT PCR_4
novel "cDNAc":	MCC RT PCR_1 & MCC RT PCR_4
novel "cDNAd":	MCC RT PCR_1 & MCC RT PCR_3

The two novel cDNAs each include an exon of CG18496 (and a short bit of genomic sequence upstream of that) which was predicted by the BDGP to be a distinct gene upstream of CG6156 (fig. 3.7). Using the primers designed for the RT PCR, cDNAc and cDNAd are predicted to be 3292bp and 1757bp long, respectively (appendix 1). It is possible that the mRNAs corresponding to these cDNAs are longer at the 5' end than cDNAc/d, although the primers used were based on the 5' terminus as predicted previously (V. Finnerty, personal communication).

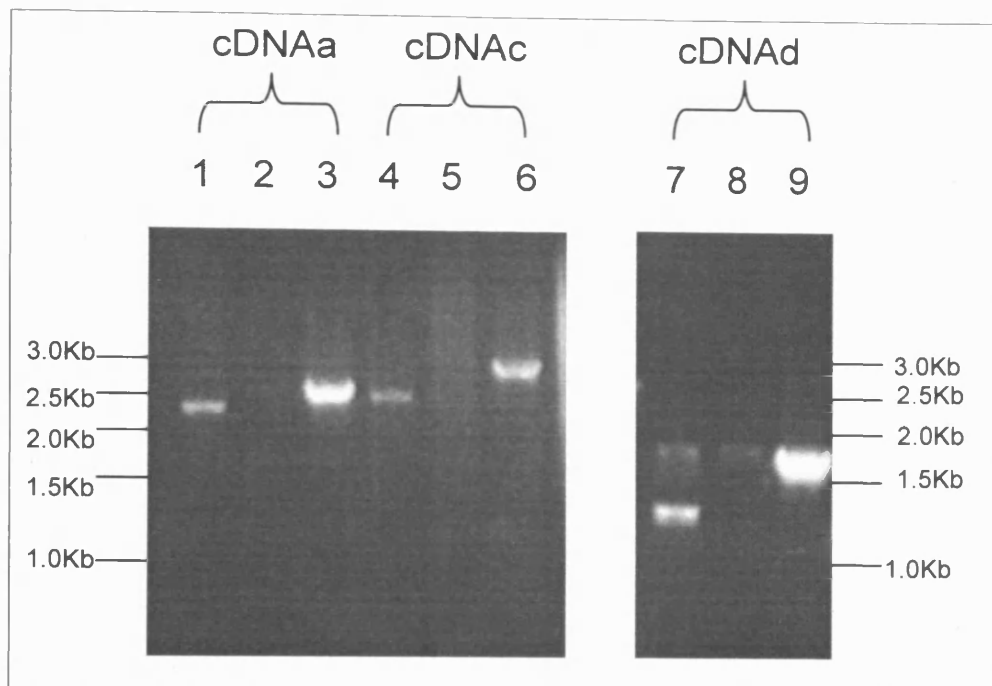


Figure 3.6. Additional MCC mRNA isoforms. cDNAs were generated by reverse transcription using RNA isolated from wild-type embryos followed by PCR using primers illustrated in **figure 3.7**. The first lane of each triplet (1, 4, 7) has reverse transcriptase present in the reaction, the middle lane (2, 5, 8) has no reverse transcriptase (control), and the final lane is a PCR using genomic DNA as template (3, 6, 9). **1-3:** Primers MCC RT PCR_2 & MCC RT PCR_4, **4-6:** Primers MCC RT PCR_1 & MCC RT PCR_4, **7-9:** Primers MCC RT PCR_1 & MCC RT PCR_3.

Since 5' RACE was not performed here for mRNAa or mRNAb, it is possible that they actually extend back to include exon A (the indicated start site in **fig.3.7** was based on sequencing of the cDNAs obtained from the BDGP). If they were to include exon A, mRNAa would share a start translation site with mRNAc. For mRNAb the start translation site would remain unchanged due to the stop codon in the read-through from exon A to B. However, these mRNAs (and therefore proteins) would still be distinct from mRNAc since they include exon B.

The mRNA corresponding to cDNAd is predicted to encode a truncated protein due to a lack of removal of the intron between exons E and F

which contains a stop codon (fig. 3.7). This protein is predicted to contain three coiled coils and the conserved arginine residue, but to lack the C-terminal two coiled coils and the PDZ target motif at its C-terminus. The RT PCR primers were designed just 3' of the translation stop codon, therefore the *in vivo* mRNA corresponding to cDNAd is likely to be significantly longer and include a 3' untranslated region. All the PCR products were confirmed by sequencing. From this point onwards, the predicted proteins encoded by mRNAc and mRNAAd will be referred to as DMCCc and DMCCd respectively.

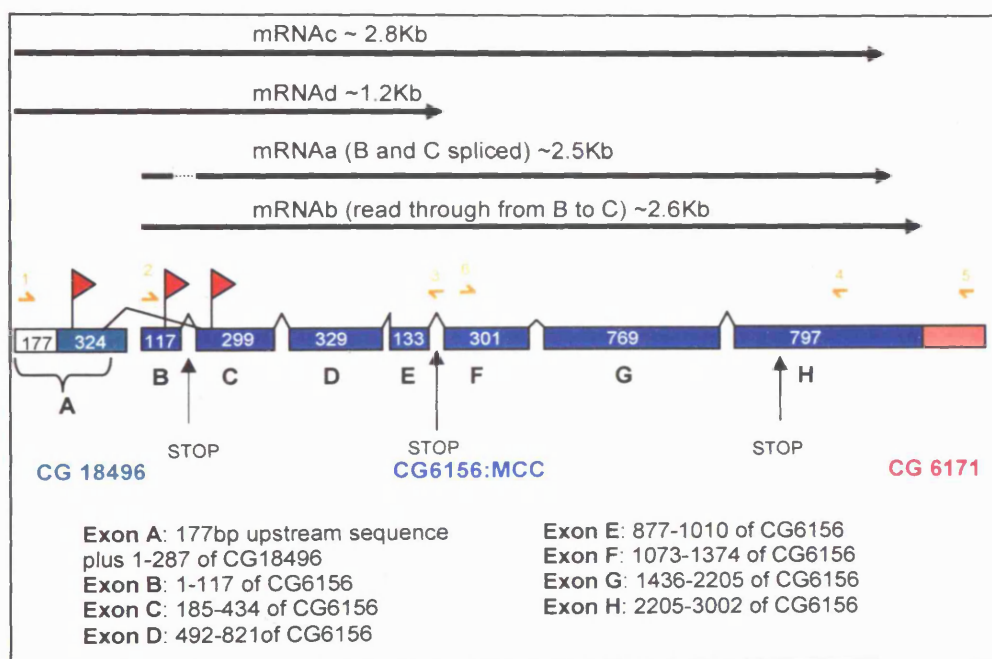


Figure 3.7. DMCC mRNAs. There are likely to be at least 4 mRNAs transcribed from the MCC gene locus. Boxes and letters denote exons (sizes in bp inside). Red flags denote methionine residues (ATG's) and STOP arrows denote stop codons. Orange arrows indicate MCC RT PCR primer positions. Not to scale.

Further RT PCRs were carried out to determine whether other mRNAs were being transcribed (fig. 3.8). Using primers MCC RT PCR_2 and MCC RT PCR_3 it was possible to show that there is unlikely to be expression of a truncated mRNA that does not remove the intron between

exons E and F and that begins at the same locus as cDNAa and cDNAb, since the cDNA corresponding to such an mRNA was not generated by RT PCR (lanes 4-6, **fig. 3.8**). Although there is PCR product in all three lanes, this is due to genomic DNA contamination since it is the size of the genomic PCR product and is present in the lane with no reverse transcriptase present in the reaction.

An RT PCR to test whether MCC mRNAs run into the downstream predicted gene (CG6171) was carried out using primers MCC RT PCR_6 and MCC RT PCR_5 (lanes 1-3, **fig. 3.8**). This also proved to be negative, indicating that the 3' end of the CG6156 locus predicted in the *Drosophila* genome annotation is likely to be correct.

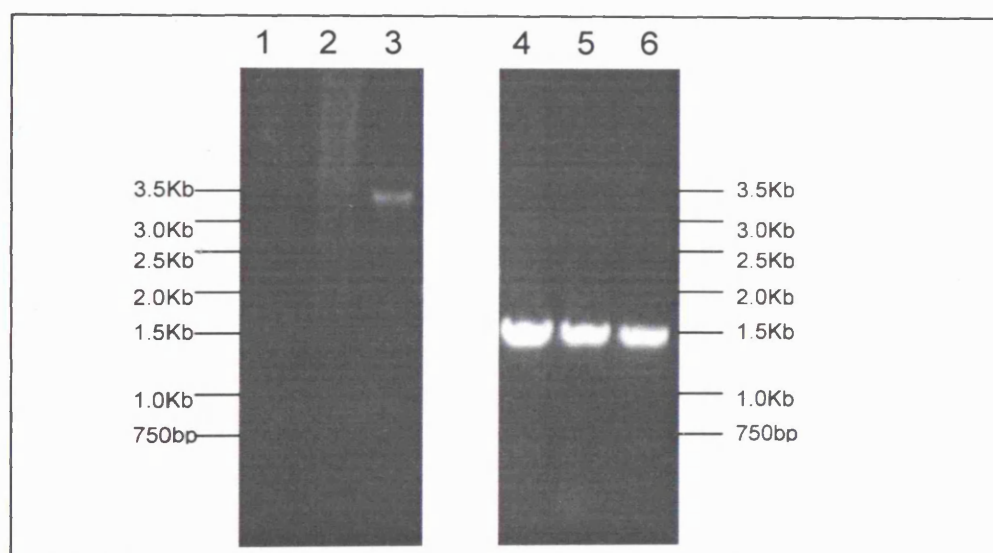


Figure 3.8. Lack of expression of further MCC mRNAs. cDNAs were generated by reverse transcription using RNA isolated from embryos followed by PCR using primers illustrated in **figure 3.7**. The first lane of each triplet (1, 4) has reverse transcriptase present in the reaction, the middle lane (2, 5) has no reverse transcriptase, and the final lane is a normal PCR using genomic DNA as template (3, 6). **1-3:** Primers MCC RT PCR_6 & MCC RT PCR_5. **4-6:** Primers MCC RT PCR_2 & MCC RT PCR_3.

3.8. DMCCc has five predicted coiled coils

DMCCc has the same five predicted coiled coils as DMCCa/b since their sequences are identical from exon C onwards (fig. 3.7). The short section encoded by the additional N-terminal exon of DMCCc and DMCCd does not contain any predicted functional domains.

3.9. All four mRNAs are transcribed during embryogenesis

The RT PCR analysis of MCC mRNAs used RNA extracted from mixed embryos at all stages of development (an overnight embryo collection). It must, therefore, be the case that all four mRNAs are transcribed at some time during embryogenesis, although it is not possible to ascertain from the RT PCR at what stage of embryogenesis they are expressed.

3.10. MCC mRNA is expressed in the developing embryonic central nervous system

In order to gain clues as to the potential function of MCC an *in situ* analysis of MCC mRNA expression during *Drosophila* embryogenesis was carried out. The probe generated for the *in situ* used primers "*in situ_1*" and "*in situ_2*" (appendix 2). This probe covers 201 to 2172bp of mRNAa and should recognise all four mRNA isoforms (fig. 3.9).

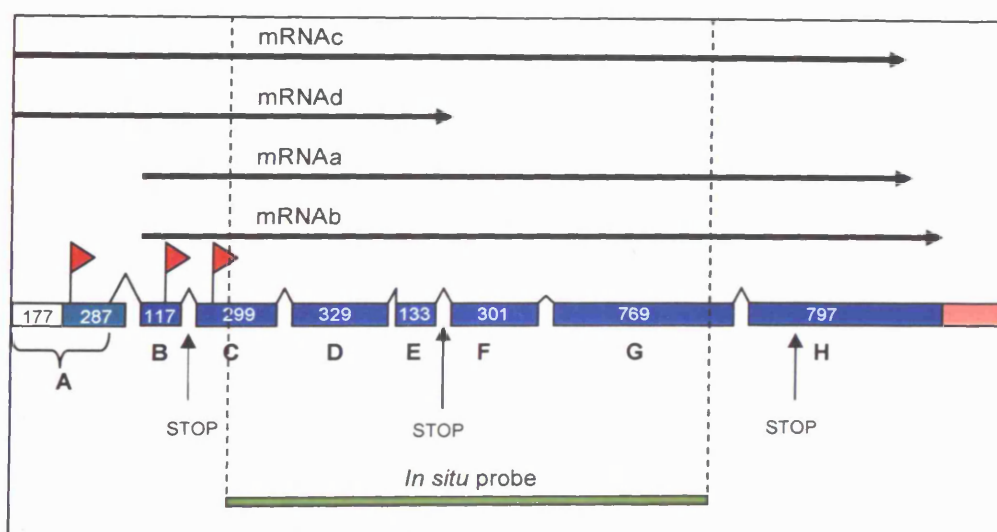


Figure 3.9. The *in situ* probe covers all four mRNAs.

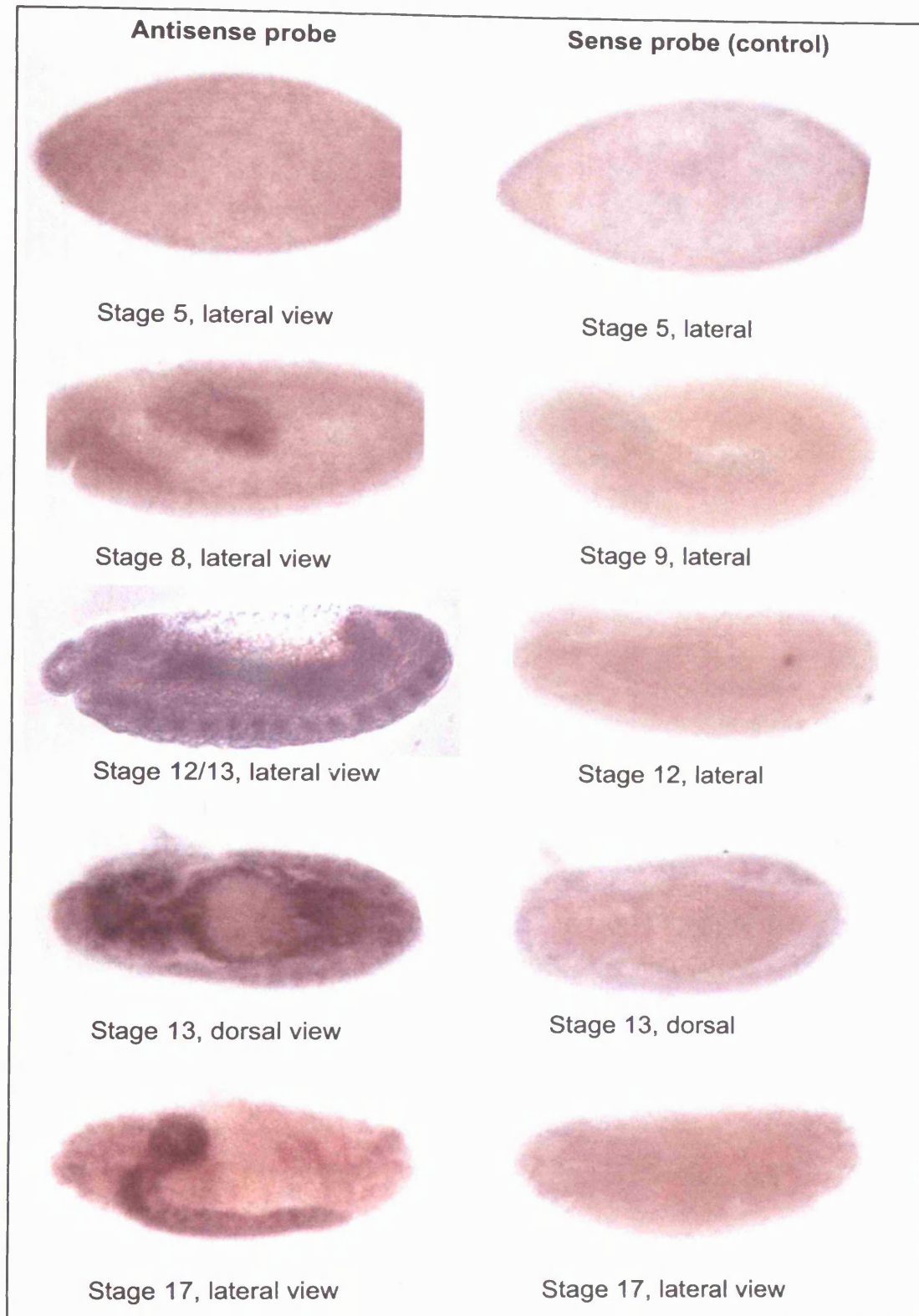


Figure 3.10. MCC mRNA expression. An overnight collection of embryos was used for *in situ* analysis. Sense (control) and antisense DIG-labelled RNA probes and standard staining procedures were used (see Materials and Methods). Anterior is to the left.

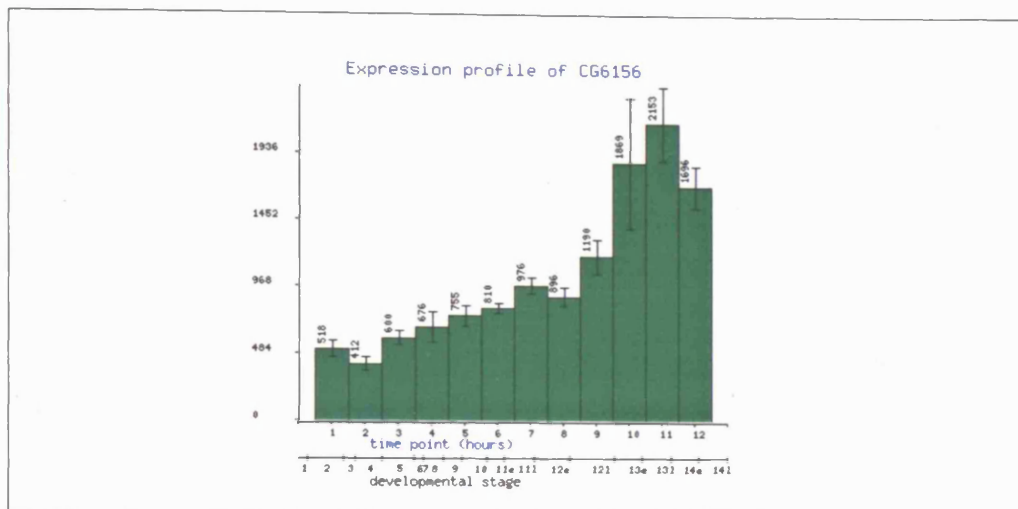


Figure 3.11. BDGP embryonic microarray expression data for CG6156. Total RNA was isolated from embryos and labeled cRNA samples were prepared following standard Affymetrix protocol. 36 Drosophila GeneChips were hybridized and scanned with Affymetrix equipment. Scanned array images were analyzed using Affymetrix and dChip analysis software. Y axis units are arbitrary. The color of the bar indicates the result of Affymetrix absent/present call that attempts to estimate whether a given gene is or is not expressed in the interrogated sample. Green indicates present (expressed), blue indicates marginal, red indicates absent. Taken from the BDGP website.

There is very low or no mRNA expression seen during stages 1-11 of embryogenesis (fig. 3.10). This fits with microarray expression data from the BDGP (fig. 3.11). From stage 12 onwards expression is observed in what appears to be a general epidermal staining pattern with high expression in the central nervous system (CNS), and possibly the anterior midgut, foregut and posterior midgut at stage 13 (fig. 3.12). Expression then becomes restricted to the central nervous system and the larval brain hemisphere from stage 14 onwards until the embryo hatches (fig. 3.13). There is no expression observed in the peripheral nervous system (PNS) or the salivary gland.

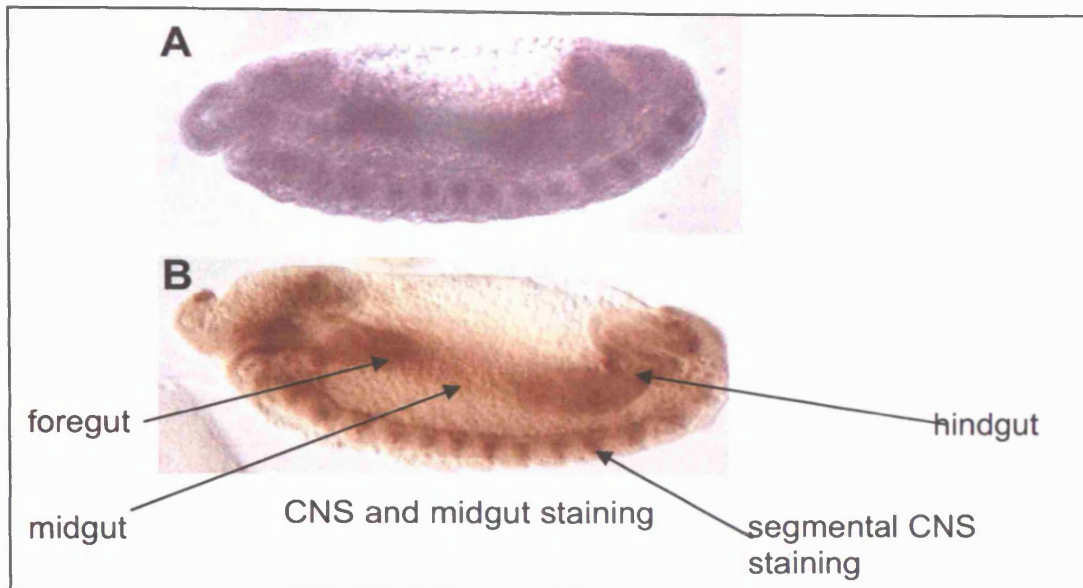


Figure 3.12. MCC may be expressed in the embryonic gut. **A** is a stage 12/13 embryo showing MCC *in situ* staining using the antisense probe. **B** is a reporter antibody staining of a stage 13 embryo from an enhancer trap line, I(2)01381, exhibiting CNS, foregut, midgut and hindgut expression (BDGP data taken from FlyView - <http://pbio07.uni-muenster.de/FlyView/Home.html>) to compare with A.

Closer analysis of the central nervous system staining reveals a segmental “spotty” pattern that is strongest on both sides of the midline (although there may also be a regular, fainter, spotty staining at the midline) (**fig. 3.13**).

According to microarray time course data, CG6156 expression decreases during the larval stages but increases once again during metamorphosis, and persists in the adult fly (**fig. 3.14**). Expression is higher in adult males than females.

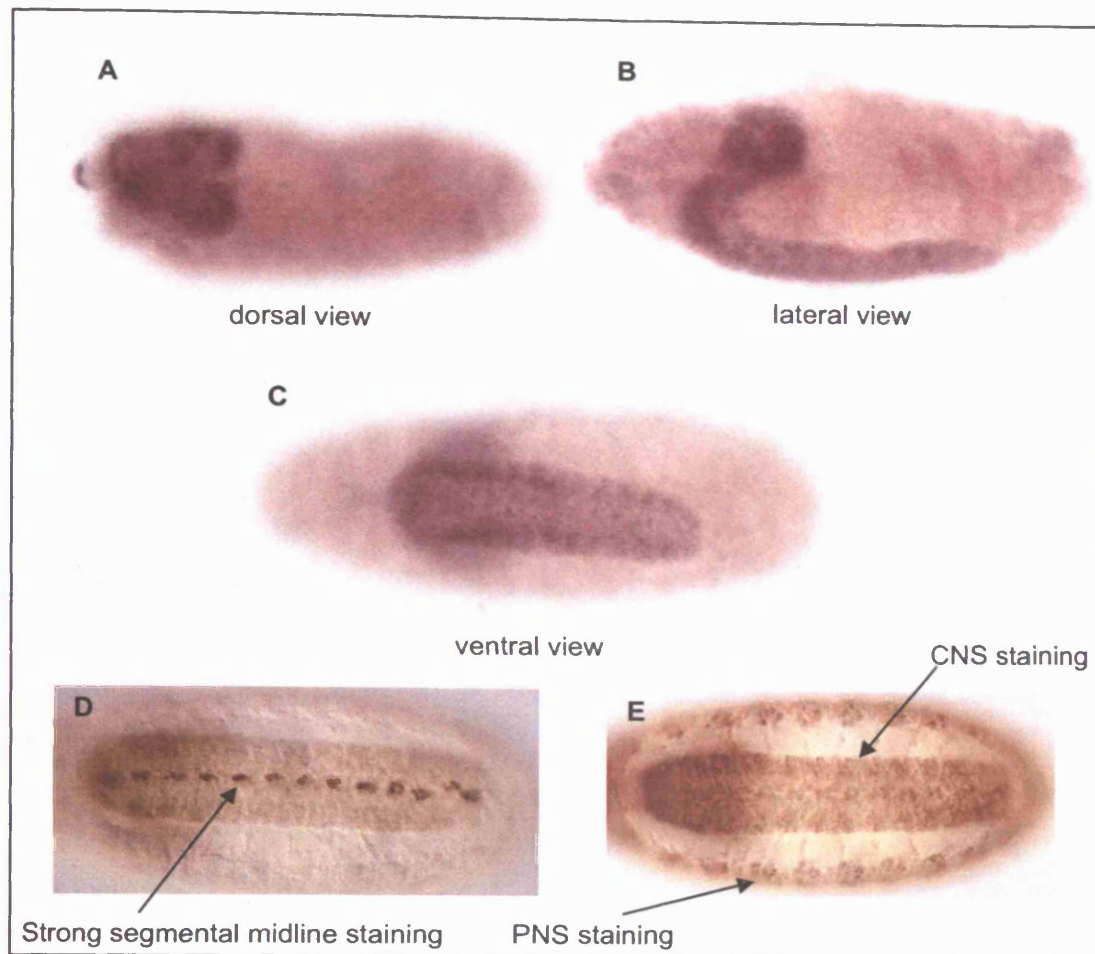


Figure 3.13. MCC expression in the central nervous system. A, B and C are stage 17 embryos showing MCC *in situ* staining. D and E are antibody staining of similar stage embryos from enhancer trap lines 10942 (D) and l(2)09327 (E) exhibiting strong midline staining and CNS staining (D) and peripheral nervous system and CNS staining (E) (taken from FlyView - <http://pbio07.uni-muenster.de/FlyView/Home.html>) to compare with C.

3.11. Discussion and conclusions

3.11.1. MCC gene and protein structure

The MCC protein is conserved through evolution from the ancestral line that led to both *Drosophila* and humans. The interaction between MCC and DRhoGEF2 via the PDZ domain may be conserved in the fruit fly, rat, fowl, chimpanzee and human. It is also possible that further species whose genomes are currently incompletely sequenced have a conserved interaction between these two proteins. Although it is currently very difficult (or at the least time-consuming) to test the interaction between

MCC and Rho GEF for all these species, it is likely that the interaction is conserved across all species since two species widely separated in evolution – fruit fly and human – conserve the interaction. The implication of this conservation of interaction during evolution is that it is important for species survival from fruit flies to humans.

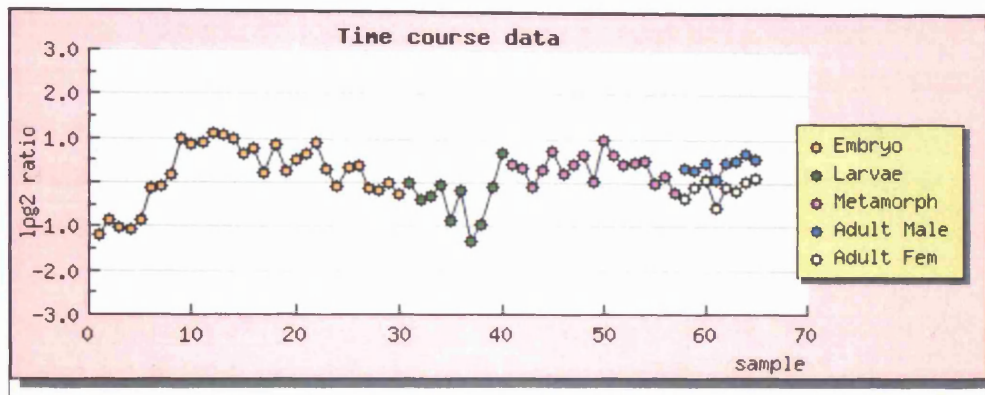


Figure 3.14. CG6156 expression from embryo to adult. Microarray time course for CG6156 expression during development. The Y axis represents arbitrary units. The X axis effectively represents time but is not a linear scale. Data taken from: flygenome.yale.edu – see (Arbeitman, Furlong et al. 2002) for methods.

It is interesting to note that an arginine residue that is frequently mutated in human patients with colorectal cancer is conserved in fruit flies and humans. This residue is thought to be required for MCC-mediated cell cycle regulation since its mutation to an alanine residue leads to an inability of MCC to block the cell cycle between the G1 and S phases (Matsumine, Senda et al. 1996).

Analysis of the primary sequence of the DMCC gene indicates that the protein shares the same predicted five coiled coil structure as the human MCC1 protein. Coiled coils are formed from bundles of α -helices wound into a superhelix. They mediate subunit oligomerisation of a large number of proteins. This can be either homo- or hetero-oligomerisation, and it is possible that MCC interacts with itself or something else via one or more of these coiled coils.

Coiled coils can exhibit a number of important properties from extreme thermostability to a means of dynamic protein folding in response to signalling (Burkhard, Stetefeld et al. 2001). MCC's coiled coils are of the short heptad-repeat variety, which are found in a variety of transcription factors and some signalling proteins. For example, the NIMA/Nek kinase-like regulator of chromosome condensation-like 1 (NERCC1) protein contains a coiled coil that mediates homodimerisation and is required for autophosphorylation (Roig, Mikhailov et al. 2002). Perhaps further study of the nature of the coiled coils of MCC, and whether they do indeed mediate oligomerisation, will reveal a function relevant to DRhoGEF2-mediated signalling.

3.11.2. MCC mRNA expression

There are at least four different mRNAs encoded by the MCC gene locus, and two of these include part of an upstream exon of what is predicted to be a distinct gene. It is possible that these four mRNAs are expressed in different tissues of the developing embryo and they may perform different or overlapping functions. The *in situ* analysis here did not distinguish between these different mRNAs since the probe is able to recognise all four. It would be interesting to use smaller probes specific to one or more of these mRNAs to determine whether they do exhibit different expression patterns. However, since the probe recognised all four mRNAs it is unlikely that any is expressed beyond the pattern revealed in **figure 3.10**.

The *in situ* analysis indicates that there is undetectable MCC mRNA expression in the early stages of embryogenesis such as gastrulation, when we know that DRhoGEF2 is expressed and is functioning. It is possible that DMCC protein is maternally-loaded, and this would not be visualised by the *in situ* analysis. An antibody against DMCC would be required to determine whether there is DMCC maternal protein loading. There is no *in situ* staining detected in the salivary glands, another tissue where DRhoGEF2 functions to promote epithelial folding (Nikolaidou and

Barrett 2004). This would imply that DMCC probably does not function in the G-protein mediated signalling pathway involving DRhoGEF2 which regulates cell shape changes and epithelial sheet folding.

DRhoGEF2 protein is expressed in the CNS during embryogenesis (Padash Barmchi, Rogers et al. 2005), therefore expression will overlap with DMCC at the later stages of embryogenesis. 8% of *Drosophila* embryos with one DRhoGEF2 null allele and one dominant negative allele have mild commissural defects in the central nervous system (W. Wei Tee and K. Nikolaidou, personal communication). Homozygous null embryos hatch into larvae that exhibit reduced locomotion during the larval stages and only 27% survive to adulthood (K. Nikolaidou, personal communication). The musculature of these larvae appears to be normal (K. Nikolaidou, personal communication), and it is possible that they have a nervous system defect leading to the “sluggish” larval phenotype. It is possible that DMCC and DRhoGEF2 function together in nervous system development.

The fact that DMCC is expressed in the CNS from stage 12/13 may indicate an involvement in neuronal differentiation since it is at stage 13 that neurons begin to differentiate and set up the classic tram-track pattern of axon fascicles (fig. 3.15). The strongest DMCC staining appears to be outside this tram-track, and is therefore likely to be in the ventral nerve cord where the ganglion mother cells and neurons reside (fig. 3.15). It is of interest to note that PDZRhoGEF and LARG, human orthologues of DRhoGEF2, are both expressed in the brain and spinal cord of mice (Kuner, Swiercz et al. 2002). It is also interesting to note that MCC is postulated to be involved in cellular differentiation in mice based on its *in vivo* expression pattern and increased expression in PC12 cells in culture in response to nerve growth factor (NGF) treatment to induce differentiation (Senda, Matsumine et al. 1999). There is currently no evidence for DRhoGEF2 involvement in neuronal differentiation.

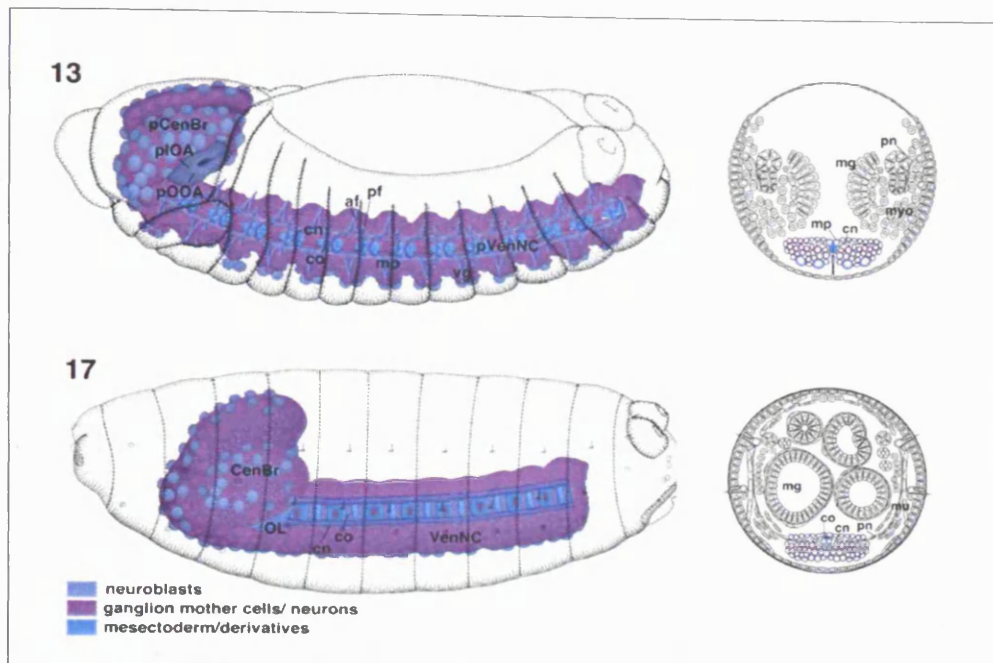


Figure 3.15. The development of the central nervous system. CenBr: brain, pIOA/pOOA: inner/outer optic anlage, cn: connectives, co: commissures, VenNC/vg: ventral nerve cord, mp: midline glial cells, OL: optic lobe, af: anterior fascicle, pf: posterior fascicle, mg: midgut, myo: myoblasts, pn: peripheral nerve, tp: tracheal pit. Dorsal is up, anterior is left. Numbers represent embryonic stages. Taken from (Hartenstein 1993)

It will be necessary to use an antibody against MCC protein in order to determine the exact nature of the CNS cells expressing MCC. A polyclonal antibody has been generated, but it proved to be non-specific both in embryo staining (no CNS staining but the yolk stained strongly, data not shown) and Western blot analysis (various dilutions of antibody were tested on S2 cells over-expressing MCC, a large number of bands were visible but none corresponded to MCC, data not shown).

3.12. Chapter 3 References

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

- Arbeitman, M. N., E. E. Furlong, et al. (2002). "Gene expression during the life cycle of *Drosophila melanogaster*." Science **297**(5590): 2270-5.
- Burkhard, P., J. Stetefeld, et al. (2001). "Coiled coils: a highly versatile protein folding motif." Trends Cell Biol **11**(2): 82-8.
- Hartenstein, V. (1993). Atlas of *Drosophila* Development, Cold Spring Harbor Laboratory Press.
- Hirano, M. and T. Hirano (2002). "Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA." Embo J **21**(21): 5733-44.
- Kinzler, K. W., M. C. Nilbert, et al. (1991). "Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers." Science **251**(4999): 1366-70.
- Kuner, R., J. M. Swiercz, et al. (2002). "Characterization of the expression of PDZ-RhoGEF, LARG and G(alpha)12/G(alpha)13 proteins in the murine nervous system." Eur J Neurosci **16**(12): 2333-41.
- Letunic, I., R. R. Copley, et al. (2004). "SMART 4.0: towards genomic data integration." Nucleic Acids Res **32 Database issue**: D142-4.
- Matsumine, A., T. Senda, et al. (1996). "MCC, a cytoplasmic protein that blocks cell cycle progression from the G0/G1 to S phase." J Biol Chem **271**(17): 10341-6.
- Nikolaidou, K. K. and K. Barrett (2004). "A Rho GTPase Signaling Pathway Is Used Reiteratively in Epithelial Folding and Potentially Selects the Outcome of Rho Activation." Curr Biol **14**(20): 1822-6.
- Padash Barmchi, M., S. Rogers, et al. (2005). "DRhoGEF2 regulates actin organization and contractility in the *Drosophila* blastoderm embryo." J Cell Biol.
- Roig, J., A. Mikhailov, et al. (2002). "Nercc1, a mammalian NIMA-family kinase, binds the Ran GTPase and regulates mitotic progression." Genes Dev **16**(13): 1640-58.

Senda, T., A. Matsumine, et al. (1999). "Localization of MCC (mutated in colorectal cancer) in various tissues of mice and its involvement in cell differentiation." J Histochem Cytochem **47**(9): 1149-58.

4. Results – DMCC functional analysis

4.1. Introduction

The manipulation of expression levels has been used extensively in *Drosophila* and other organisms to study the function of genes. The artificial increase or decrease (or complete knock out) of expression of a gene will normally lead to changes in mRNA and protein levels, and in many cases lead to a phenotype that can be visualised and characterised. With the vast array of *Drosophila* phenotypic data available, often the phenotype of a mutant can be used to place the gene in a specific signalling pathway or at least to ascertain that it is involved in a specific developmental process.

In this chapter the function of the MCC gene is investigated further by the manipulation of mRNA expression levels, generation of an MCC mutant, and phenotypic analysis. mRNA levels are manipulated by both over-expression using the GAL4-UAS system, and knock down of expression using RNAi. The mutant is generated by the hopping of a P-element, identified by PCR, and confirmed by Southern hybridisation.

4.2. The GAL4-UAS system

The GAL4-UAS system makes use of a yeast transcription activating factor, GAL4, and the DNA sequence to which it binds, Upstream Activating Sequence (UAS), to regulate gene expression. This system can be used in *Drosophila* to ectopically express a gene of interest in a tissue of interest (Phelps and Brand 1998). The gene of interest is fused downstream of a UAS and this is crossed to a fly carrying a GAL4 gene under the control of a promoter or enhancer. This gives rise to expression of the gene of interest in a specific tissue or set of tissues as determined by the promoter driving GAL4 expression. Large numbers of *Drosophila* GAL4 lines have been generated ("GAL4 drivers") to enable expression of a gene in almost any tissue desired.

4.3. Over-expression of MCC using the GAL4-UAS system

The GAL4-UAS system was used to ectopically express the MCC gene in various tissues. The GAL4 drivers used are described in **table 2.1** of Materials and Methods. Three different UAS-MCC constructs were used for over-expression (gratefully received from V. Finnerty, Emory University, GA): UAS-DMCCc, UAS-DMCCd and UAS-MCC1.

4.4. Over-expression of MCC leads to no obvious developmental phenotype

Since MCC is expressed in the central nervous system, the elav-GAL4 driver was used to over-express both DMCC and HsMCC specifically in nervous tissues. Over-expression of MCC in nervous tissues gives rise to viable flies with no visible phenotype (**fig. 4.1**). In order to test whether ectopic expression of MCC produces a phenotype, over-expression of MCC specifically in the eye using the sevenless-GAL4 or eyeless-GAL4 drivers was carried out, or in the wing using the wingless-GAL4 driver (not shown). Ectopic expression using these drivers also results in no obviously visible phenotype (**fig. 4.2**).

Over-expression in the eye using the Glass Multimer Reporter (GMR)-GAL4 driver results in flies with rough eyes due to disruption of the ommatidial architecture. However, GMR-GAL4 flies themselves have rough eyes so the effect is not specific to MCC over-expression (**fig. 4.2**). The eyes did not appear to be more or less rough when MCC was over-expressed.

In order to test whether ectopic expression of MCC in any tissue gives rise to a phenotype, MCC was ubiquitously over-expressed using the tubulin-GAL4 driver. This driver over-expressing MCC has no effect on viability (**table 4.1**) and gives no obviously visible phenotype (**fig. 4.1**).

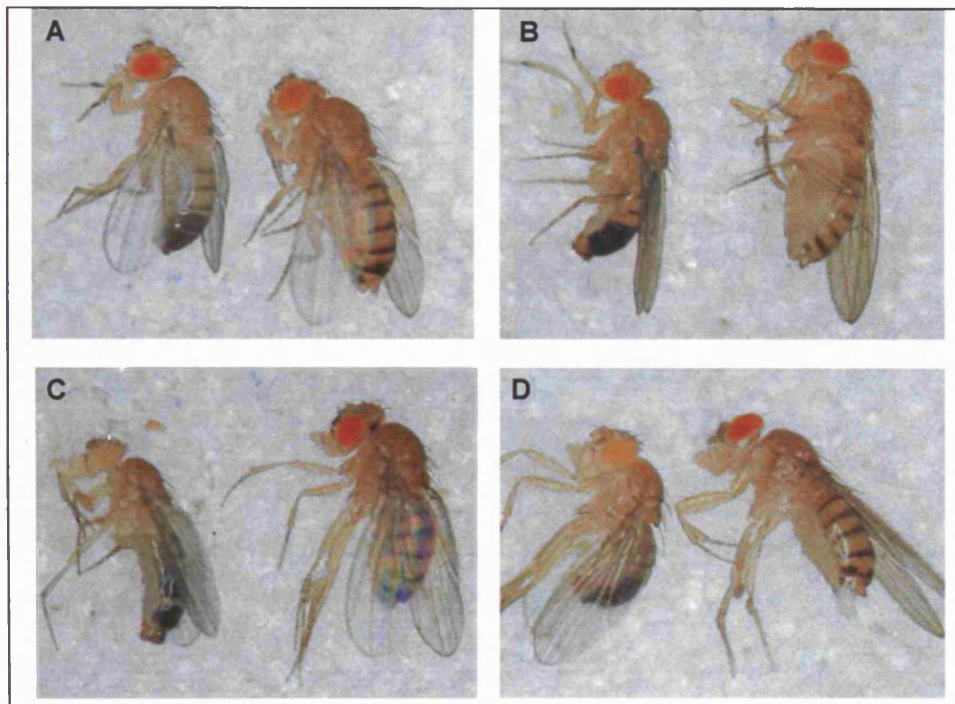


Figure 4.1. Over-expression of DMCC and MCC1 in *Drosophila*.

Male (left) and female flies resulting from crosses to cause over-expression of MCC. **A:** Tubulin GAL4 x UAS-DMCC, **B:** Tubulin GAL4 x UAS-MCC1, **C:** Elav-GAL4 x UAS-DMCCc, **D:** Elav-GAL4 x UAS-MCC1. Over-expression of DMCCd is not shown but flies looked identical to DMCCc.

4.5. The use of RNA interference to compromise MCC function

RNA interference (RNAi) technology is used to compromise gene expression in order to study gene function. dsRNA corresponding to a gene of interest is introduced into the system provoking an RNAi response that degrades the dsRNA and leads to the additional specific destruction of mRNA for that particular gene. The use of RNAi was first shown in the worm (Fire, Xu et al. 1998). The exact molecular nature of this response is not entirely understood, but it has been shown to exist in various systems including both *Drosophila* and mammalian cells in culture, and in *Drosophila* embryos (Hannon 2002).

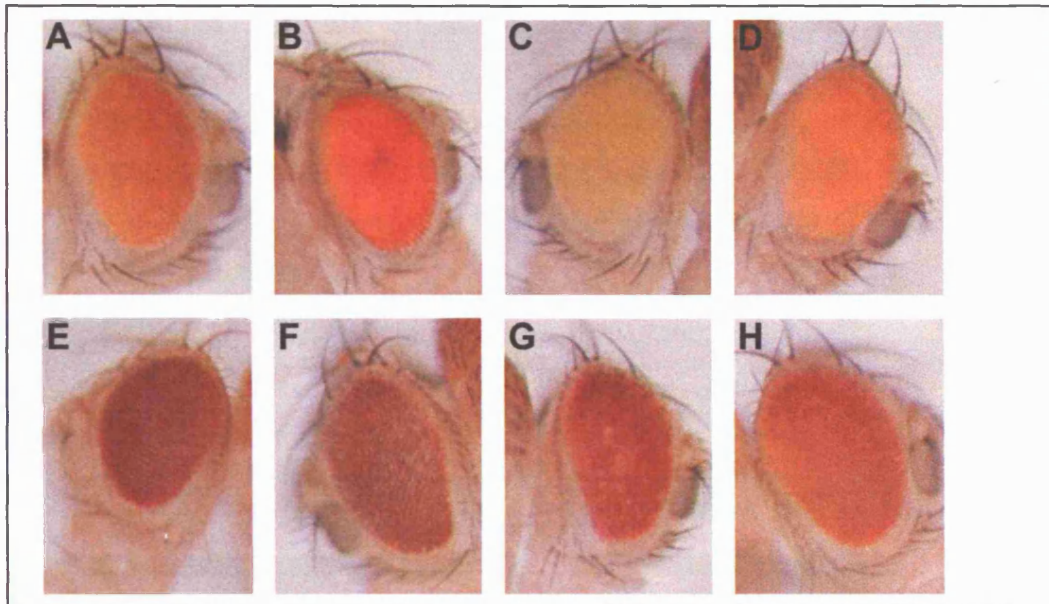


Figure 4.2. Over-expression of MCC specifically in the *Drosophila* eye.

The UAS-MCC constructs were over-expressed specifically in the eye using the Sev-GAL4 or GMR-GAL4 driver. **A:** sevGAL4 x UAS-DMCCc, **B:** sev-GAL4 x UAS-DMCCd, **C:** ey-GAL4 x UAS-DMCCc, **D:** ey-GAL4 x UAS-DMCCd, **E:** GMR-GAL4, **F:** GMR-GAL4 x UAS-DMCCc, **G:** GMR-GAL4 x UAS-DMCCd, **H:** GMR-GAL4 x UAS-MCC1.

	Tub-GAL4	TM3, Sb
UAS-DMCCc	(116) 49%	(120) 51%
UAS-DMCCd	(126) 50%	(123) 50%
UAS-HsMCC	(145) 52%	(135) 48%

Table 4.1. Ubiquitous over-expression of MCC. *Drosophila* crosses were performed at 25°C and emerging adults scored from 10 days onwards until all had emerged. Percentages are the percentage number of flies of each genotype resulting from the cross. Numbers in parentheses are actual number of adult flies.

4.6. Generation of a UAS-MCC RNAi DNA construct

An RNAi construct was generated as a “flipback” cDNA:genomic DNA hybrid (Kalidas and Smith 2002). An earlier attempt was made to generate a cDNA:cDNA flipback construct, but it was not possible to sub-clone this using bacteria probably due to the secondary structure formed

by the DNA. The advantage of the genomic DNA:cDNA hybrid is that, at the DNA level, there is less opportunity for secondary structure formation, but, once splicing has occurred to produce the mature mRNA, secondary structure can form.

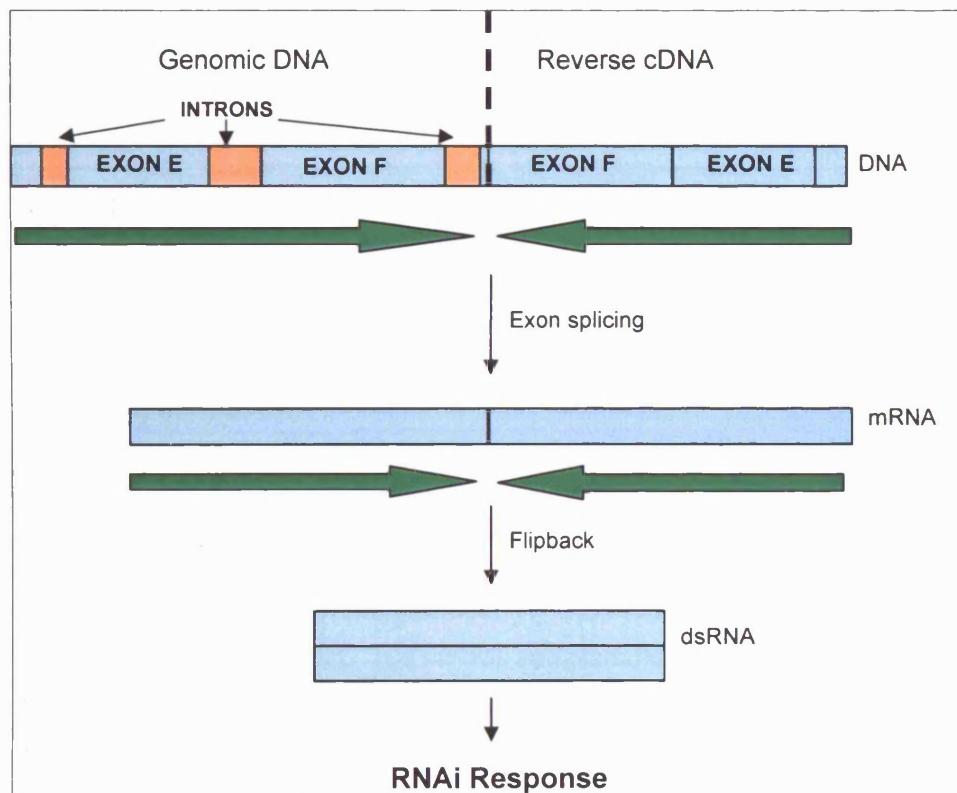


Figure 4.3. Hybrid genomic DNA:cDNA construct. Introns are removed during the mRNA maturation process and the RNA flips back on itself to create the dsRNA species that activates the RNAi response. Exon lettering as in **fig. 3.7** of chapter 3. Not to scale.

A region covering two and a half exons of CG6156 genomic sequence (from 777 to 1442bp of the published genomic sequence (**appendix 1**)) was fused to the corresponding reverse complement sequence of cDNA (**fig. 4.3**). The first five nucleotides of exon G plus a restriction site act as a linker between the genomic and cDNA pieces. Both fragments were generated by PCR (**fig. 4.4**) and sub-cloned into pUASp which was chosen since it enables expression in the germline and very early stage embryos, and can be used to express MCC-RNAi in a tissue of choice

using the GAL4-UAS system (Rorth 1998). This vector contains the elements required for P-element mediated stable insertion into the *Drosophila* genome (Rubin and Spradling 1982).

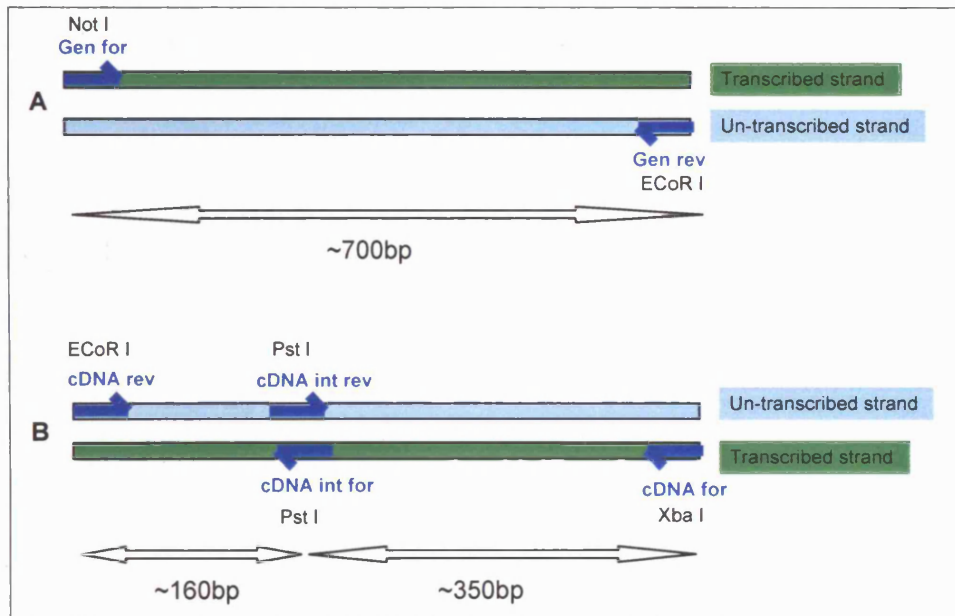


Figure 4.4. Generation of the genomic (A) and reverse cDNA (B) pieces of the RNAi hybrid construct. The reverse complement full-length mRNA sequence was scanned using DNA Strider to check for intron splice donor sites (GTnnGT) revealing many such sites. The cDNA piece of the hybrid was selected due to it encompassing only one intron donor site. This piece was made in two parts by PCR with the intron splice donor site mutated to a Pst I restriction site which was used to join the two parts together. Restriction sites were also introduced via primers (**appendix 2**) at both ends of the complete cDNA piece to enable joining to the genomic piece and ligation into the pUASp vector. The genomic piece of the hybrid was generated by PCR using primers containing restriction sites to enable joining to the cDNA piece and ligation into the pUASp vector. Blue arrows represent PCR primers. Not to scale.

4.7. Generation of transgenic UAS-MCC-RNAi flies and crossing to GAL4 drivers

~3000 early embryos were injected with the pUASp-MCC-RNAi construct and from these ~100 individual transgenic lines were obtained. These lines were mainly homozygous viable and maintained as such. A

selection was crossed to various GAL4 drivers in order to activate expression of the MCC-RNAi construct and thereby compromise MCC gene expression. The GAL4 drivers used include two that give ubiquitous expression and one that gives CNS expression. Four transgenic UAS-RNAi lines that carried insertions on different chromosomes were used:

RNAi⁶⁵, RNAi⁹⁸ – insertion(s) on 3rd chromosome

RNAi⁴⁶ - insertion(s) on 2nd chromosome

RNAi²⁹ - insertions on X and 2nd chromosomes

4.8. Flies expressing an MCC-RNAi construct are viable with no visible phenotype

Flies that expressed the RNAi construct ubiquitously under the control of either the tubulin GAL4 or daughterless-GAL4 (data not shown) drivers were viable and had no visible phenotype (**table 4.2**). Flies that strongly expressed the construct in the nervous system (under the control of the elav-GAL4 driver) were also viable with no visible phenotype (data not shown).

	Tubulin-GAL4	TM3, Sb
RNAi⁶⁵	(156) 51%	(149) 49%
RNAi⁹⁸	(136) 50%	(139) 50%
RNAi⁴⁶	(124) 51%	(119) 49%
RNAi²⁹	(148) 52%	(137) 48%

Table 4.2. MCC RNAi expression studies. *Drosophila* crosses were performed at 25°C and emerging adults scored from 10 days onwards until all had emerged. Percentages are the percentage number of flies of each genotype resulting from the cross. Numbers in parentheses are actual number of adult flies.

4.9. Strategy for generation of an MCC mutant

In parallel with generating transgenic RNAi flies, a second strategy was employed to knock out MCC gene function in *Drosophila*. It was hoped

that, by taking a two-pronged approach to gene knockdown (i.e. RNAi and mutagenesis), a phenotype could be assessed more rapidly, and useful reagents generated in the process.

The generation of mutants has been used for many years to study the function of proteins. Since the phenotype of an MCC mutant fly is unknown, and is not certain to be lethal, a reverse genetics approach was applied. Reverse genetics starts with a gene and moves towards finding a phenotype associated with that gene. The strategy for finding a phenotype associated with the MCC gene involved making a mutant using the technique of P-element hopping.

4.10. P-elements are mobile transposons in *Drosophila* that can be used for mutagenesis

P-elements are large (up to ~10Kb) pieces of DNA that are able to transpose themselves around the *Drosophila* genome inserting more or less randomly into a chromosome (O'Hare and Rubin 1983). These elements can be used for mutagenesis of *Drosophila* genes by screening for insertion of a P-element into a gene of interest (Bellen, Levis et al. 2004). The insertion of such a large piece of DNA into a gene (or upstream regulatory region) is very likely to disrupt expression of that gene, and in most cases leads to complete abolition of gene expression and a resulting null phenotype. The strategy of P-element insertion mutagenesis was chosen for MCC mutagenesis since the technique is well-characterised and there is a P-element within the vicinity of the MCC gene available.

4.11. The P-element starting point

P-elements are more likely to hop locally than to hop long distances along a chromosome or move chromosomes (Tower, Karpen et al. 1993). In order to maximise the chances of a P-element hop resulting in an insertion into the MCC gene, a fly line with a P-element as close to the MCC gene as possible was used as the P-element starting point. This

line carries a P(lacW) P-element (Bier, Vaessin et al. 1989) ~7.8Kb upstream of the CG6156 locus in the 88F1 region of the 3rd chromosome right arm (**fig. 4.5**). This P-element is not within any predicted genes but does lie in a relatively gene-rich region. This insertion is called P0 throughout this chapter.

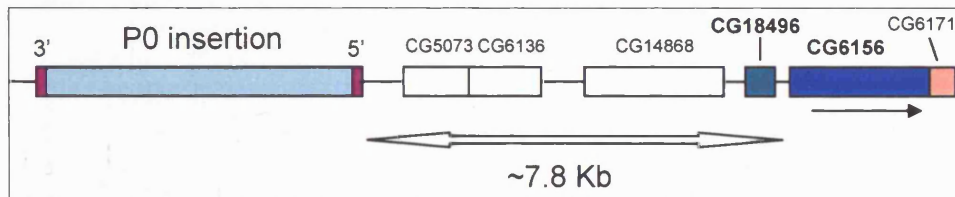


Figure 4.5. Location of the original insertion (P0). Purple boxes represent 31bp inverted terminal repeats of P-element. Not to scale.

Drosophila carrying the P0 insertion are homozygous viable with a distinctive phenotype. This phenotype is characterised by bristles on the thorax that are three-quarters of the normal length (**fig. 4.6**).

The posterior scutellar bristles are often, but not always, kinked and have the appearance of being knotted together. Since the P0 insertion is not within a predicted gene (it is between CG14867 and CG5073), it is likely that its insertion affects the expression of a gene some distance away, perhaps by affecting the function of an enhancer, resulting in the bristle phenotype.

A plasmid rescue making use of the ampicillin resistance gene within P(lacW) (**fig. 4.7**) was performed to confirm the published insertion site in the P0 line. Plasmid rescue involves restriction digestion of genomic DNA, ligation to circularise fragments and transformation into bacteria under antibiotic selection. DNA from twelve bacterial colonies were sequenced (six in each direction) and all gave the predicted insertion site indicating that there are not likely to be other P-element insertions present in this line and that the predicted insertion site is correct. The P0 line was crossed to a source of transposase, and resulting new insertion

lines were balanced for subsequent screening (see Materials and methods for details of the crosses).

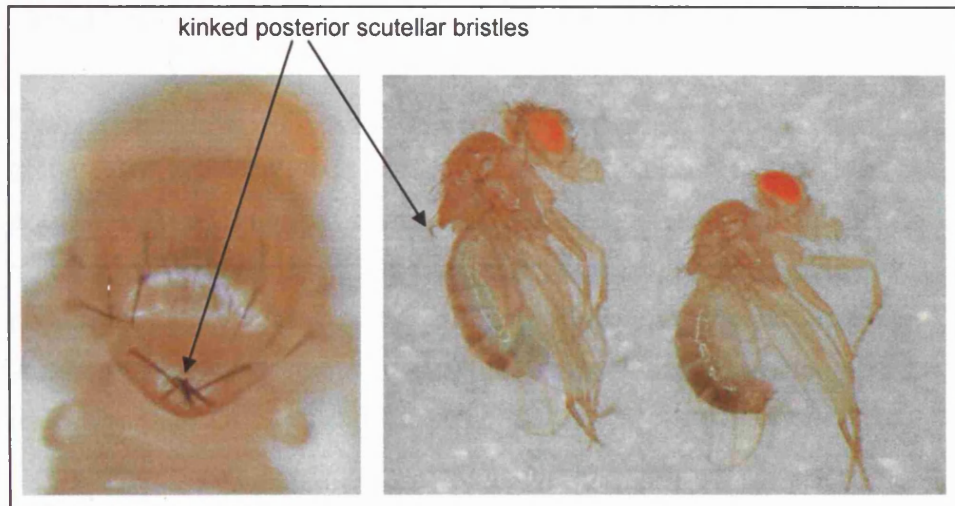


Figure 4.6. The P0 bristle phenotype. *yw* ; *P(lacW^{mw})88F1*

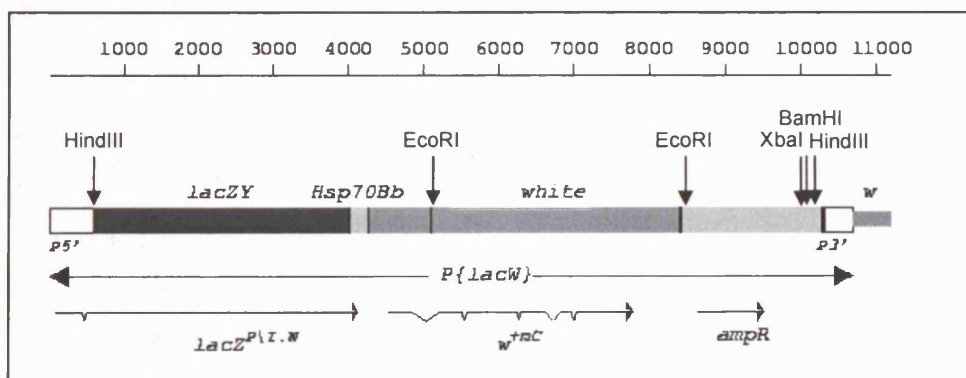


Figure 4.7. P(lacW). *lacZY*: E.coli β -galactosidase gene, *Hsp70Bb*: *Drosophila* heat shock protein 70Bb, *white*: mini-white gene, *ampR*: ampicillin resistance gene. Numbers represent base pairs. Relevant restriction sites are indicated. Taken from Flybase.

4.12. PCR screening of P-element insertions

In order that no phenotypic assumptions were made, a PCR strategy was used to screen all lines for insertions into the MCC gene. P-element primers were designed for both ends of the P-element just internal to the terminal repeats (fig. 4.8). Genomic primers were designed so that the

PCRs covered the complete CG6156 locus, the CG18496 locus and most of the upstream region that forms the start of mRNA_{Ac} and d (fig. 4.8, see appendix 2 for primer sequences). The PCRs tested both ends of the P-element in combination with all the genomic primers.

Flies were grouped into pools for PCR screening. To test the sensitivity of the PCR, one P0 fly was pooled with increasing numbers of OreR flies and a P-element/genomic primer combination used in a PCR. It was determined that one P0 fly in a pool with 19 wild-type flies could be detected by PCR, therefore flies were grouped in pools of 20. Pooling into numbers greater than this resulted in a decrease in sensitivity that made the PCR unreliable (data not shown).

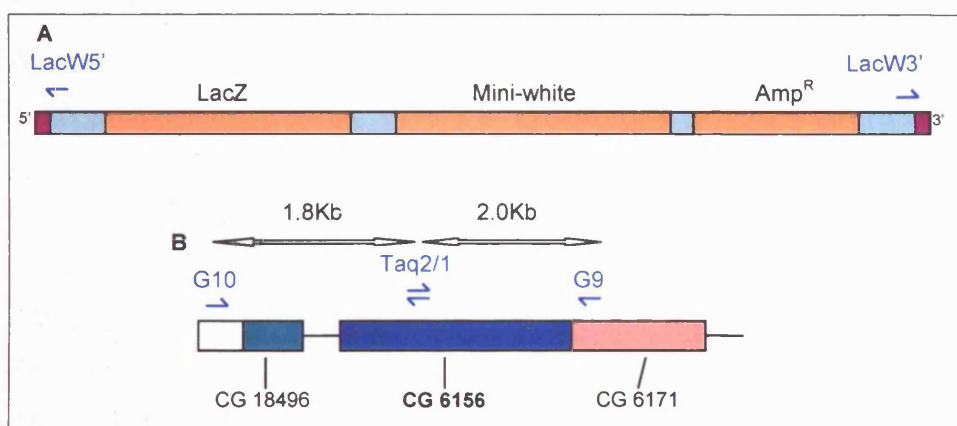


Figure 4.8. Primer positions for PCR screening. A: P(lacW) primer positions, purple boxes represent 31bp terminal repeats in P-element. B: genomic primer positions. Not to scale.

PCRs were run in 96-well format with the 12th column used exclusively for controls (table 4.3). Taq 3 and Taq 4 were additional primers designed approximately 2Kb from each end of the P0 insertion for use in control PCRs to make sure the P-element primers were working. It was assumed that at least one of the twenty flies in each pool would still have a P-element at the original insertion site in order for this control to work. The genomic DNA extracted from each pool of twenty was run on an agarose gel to check quality and concentration before being subjected to PCR.

Genomic DNA 1-20	21- 40	41- 60	61- 80	81- 100	101- 120	121- 140	141- 160	161- 180	181- 200	201- 220	Controls
G9 + lacW5'											▶ G9 + lacW5' + Taq2
G9 + lacW3'											▶ G9 + lacW3' + Taq4
G10 + lac W5'											▶ G10 + lacW5' + Taq3
G10 + lac W3'											▶ G10 + lacW3' + Taq1
Taq1 + lacW5'											▶ Taq1 + lacW5' + P10
Taq1 + lacW3'											▶ Taq1 + lacW3' + Taq4
Taq 2 + lacW5'											▶ Taq 2 + lacW5' + Taq 3
Taq2 + lacW3'											▶ Taq2 + lacW3' + G9

Table 4.3. 96-well PCR screening. The genomic DNA indicated in the top row (e.g. 1-20) was added to each of the 8 wells in the below column and thereby tested with all primer combinations. A genomic DNA prep was chosen at random for each control.

4.13. A single insertion into MCC

Approximately 1500 lines were screened by PCR. Of these only one pool of 20 gave a positive result. A band of approximately 1.4Kb was visible on the PCR screening gel in the lane corresponding to the primer pair G10 + LacW5' (fig. 4.9).

The genomic DNA from all 20 lines was extracted individually and PCRs identified one line carrying the insertion (not shown). The PCR product was sequenced and indicated an insertion within exon D, 623bp into mRNAa (fig. 4.10). This insertion (named P1) is predicted to affect the expression of all four mRNAs. Flies carrying this insertion are homozygous lethal.

4.14. P1 is not a “clean” insertion

Although the primer pair G10 + LacW5' gave a PCR product, the LacW3' primer did not give a PCR product with the Taq1 primer. A number of other genomic primers to the right of the Taq1 primer were tested in combination with the LacW3' primer, but none of these gave a PCR product (data not shown). This implies that the 3' end of the P-element and/or the 3' end of CG6156 may be deleted in this insertion line.

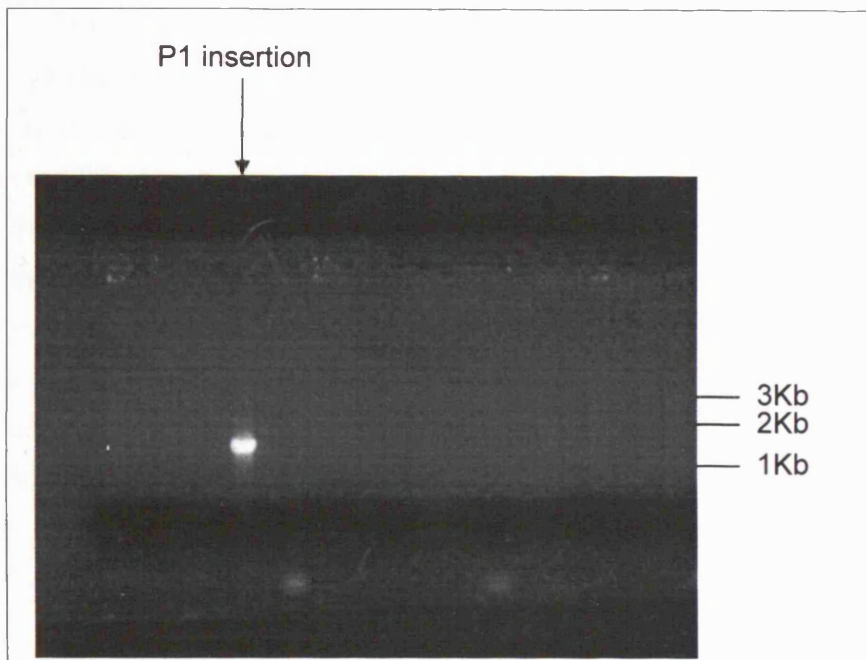


Figure 4.9. P1 insertion discovery. Genomic DNA from pools of 20 P-element hop flies was extracted and subjected to PCR using primers that covered the MCC gene. Image shows part of the screening gel for lines 641-720.

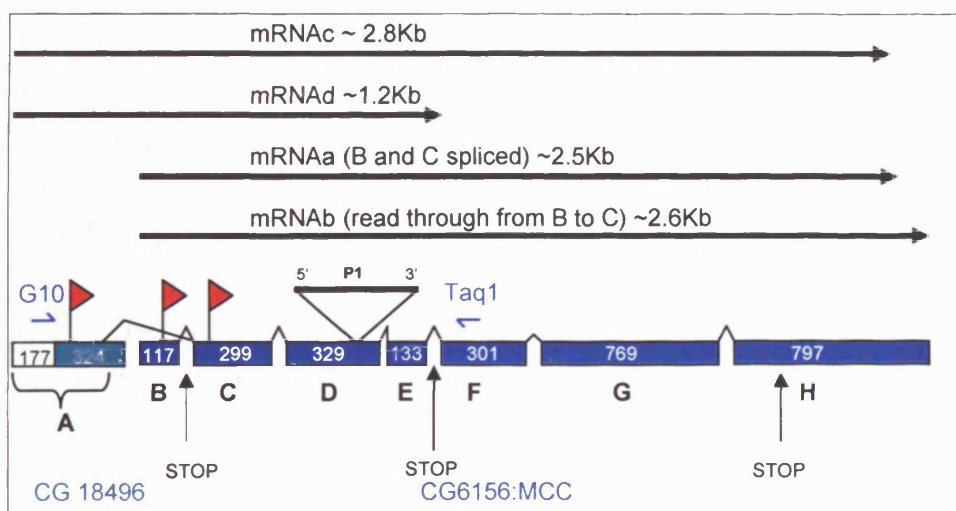


Figure 4.10. P1 insertion position. The P1 insertion is within exon D (the third exon of CG6156). Not to scale.

4.15. Molecular analysis of the P1 insertion

In order to determine the extent of any possible deletion of the P1 P-element, PCRs were performed using primers internal to the P-element in combination with the G10 primer (**fig. 4.11**). PCRs up to and including the “8.8out” primer gave PCR products of the predicted size indicating that at least 8.8Kb of the P-element is present (**fig. 4.12**). A PCR using the LacW3'probe1 primer did not give PCR product indicating that less than 10.5Kb is likely to be present (the faint band on the gel in lane 3 was likely to be loading spill-over from lane 2 since subsequent PCRs with this primer combination gave no product - personal communication, E. Batchelor).

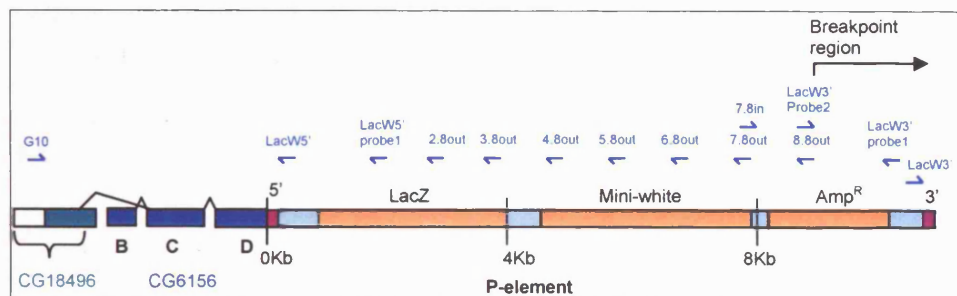


Figure 4.11. Molecular mapping of the P1 P-element deletion. Primers were designed to cover the whole P-element. Not to scale

Since more than 8.8Kb but less than 10.5Kb of the P-element is present, the breakpoint is likely to occur within the ampicillin resistance gene. Plasmid rescue experiments did not amplify the P1 insertion confirming that the ampicillin resistance gene is not intact. However, colonies were obtained, all of which corresponded to the P0 insertion, indicating that this is still present in the P1 line. This also indicates that there are unlikely to be further intact insertions in this line since all 18 colonies that were sequenced (10 in one direction, 8 in the other) corresponded to the original insertion. The P1 flies have darker eyes than the original insertion line which could be indicative of two insertions each with a mini-white gene.

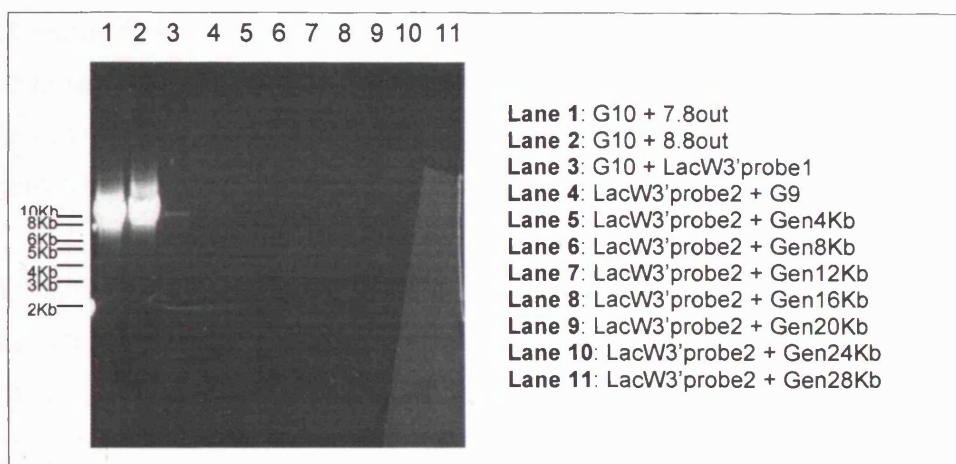


Figure 4.12. Determining the extent of deletion of the P-element in P1 and investigating a possible 3' genomic deletion. Genomic DNA was extracted from flies carrying the P1 insertion and subject to PCR using primers internal to the P-element in combination with the G10 primer (figs. 4.8 and 4.11).

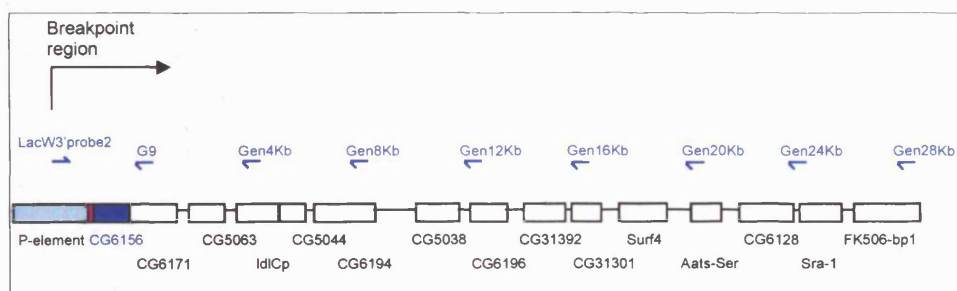


Figure 4.13. Molecular mapping of the P1 insertion line deletion. Primers were designed to cover the region up to 28Kb to the right of CG6156. Not to scale.

It is possible that the deletion event removing the 3' end of the P-element also removed some of the 3' end of CG6156 and possibly chromosome to the right of this also. To address this possibility, PCRs were performed using a "LacW3'probe2" primer that is the reverse complement of the "8.8out" primer (fig. 4.11) in combination with a number of genomic primers at the 3' end of the CG6156 locus and up to 28Kb to the right of the 3' end of CG6156 (fig. 4.13). PCRs using primers up to and including the Gen28Kb primer were all negative, indicating that up to 28Kb to the right of the 3' end of CG6156 may be deleted in addition to the 3' end of

the P-element (**fig. 4.12**). The faint band in lane 9 was not seen in subsequent PCRs with the same primer pair and can probably be attributed to non-specific PCR priming. As can be seen in **figure 4.13**, the region to the right of CG6156 is very gene (or predicted gene) –rich, therefore any deletion is likely to remove a large number of genes.

In another attempt to determine the extent of the deletion molecularly, restriction digestion of genomic DNA from the P1 line was followed by ligation to circularise fragments. The selected restriction enzymes cut within the P-element 5' of the breakpoint, but do not cut again within the P-element and would then cut somewhere within the genomic DNA to the right of the P-element (**fig. 4.14**).

Using primers internal to the P-element, PCRs were attempted across the breakpoint in an inverse PCR (e.g. for *Apal* using primers 7.8out + LacW3'probe2). However, these PCRs did not produce product (**fig. 4.15**). Since the original P0 insertion, which is still present in the P1 line, should have given PCR product and did not, it would seem that this technique was not successful.

4.16. The deletion does not extend as far as the easter gene

In order to narrow down more quickly the extent of the deletion in the P1 line, a genetic complementation approach was used. The MCC gene sits in the 88F1 region of the 3rd chromosome right arm. The *easter* gene is the nearest gene locus to the right of CG6156 for which there is a mutant available. *easter* lies approximately 33Kb downstream of the 3' end of CG6156. The P1 line was crossed to two *easter* mutant alleles (*ea*¹ and *ea*¹⁴), both of which are homozygous lethal at 25°C.

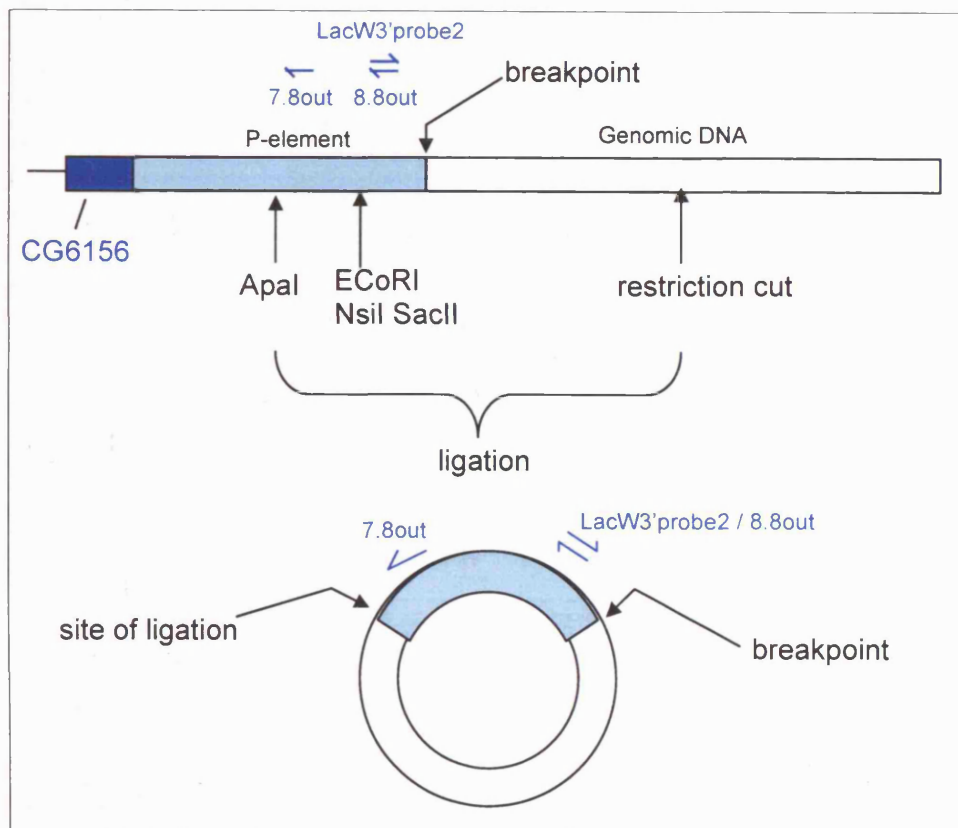


Figure 4.14. Strategy for sequencing to the right of the P1 insertion to determine the extent of deletion. Restriction digestion of genomic DNA from the P1 line was followed by ligation to circularise fragments. The selected restriction enzymes cut within the P-element 5' to the breakpoint, but do not cut again within the P-element and would then cut somewhere within the genomic DNA to the right of the P-element. The resulting circularised fragments were subject to PCR using primers internal to the P-element. ApaI, EcoRI, NsiI and SacII were all used in separate reactions.

The P1 line complements both alleles indicating that the *easter* gene is not deleted in the P1 line (table 4.4). This, combined with the PCR results, leads to the conclusion that the right-hand breakpoint of the deletion is between 28Kb and 33Kb to the right of the 3' end of CG6156.

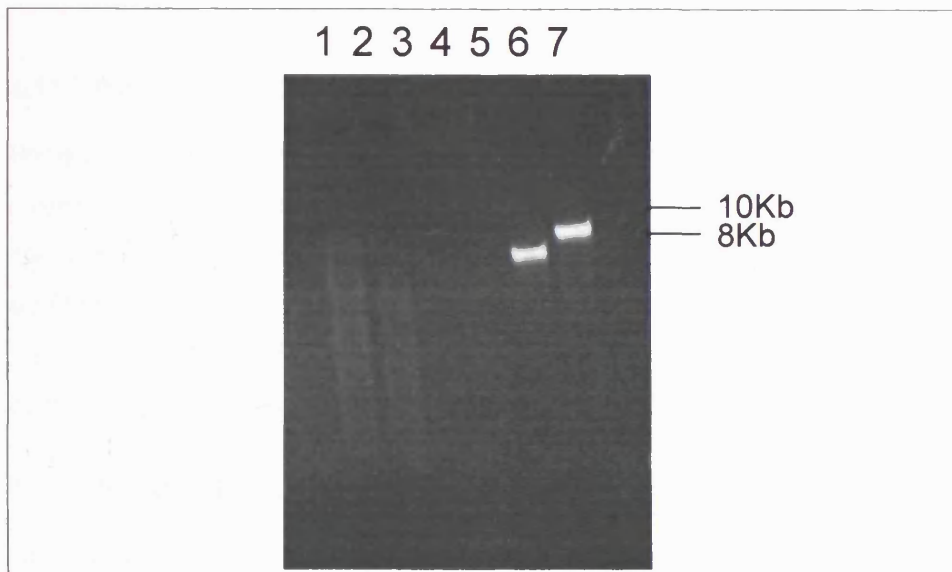


Figure 4.15. Ligation and PCR to the right of the P-element in P1 line. For methods see **fig. 4.14** legend. **1:** Apal digestion, PCR using 7.8out + LacW3'probe2 primers. **2:** ECoRI digestion, PCR using 8.8out + LacW3'probe2 primers. **3:** NsiI digestion, PCR using 8.8out + LacW3'probe2 primers. **4:** SacII digestion, PCR using 8.8out + LacW3'probe2 primers. **5:** Apal digestion, PCR using primers 4.8out + LacW5'probe2. **6:** Control PCR with P1 genomic DNA using primers G10 + 5.8out, **7:** Control PCR with P1 genomic DNA using primers G10 + 8.8out.

	<u>P1</u>	<u>TM6C, Sb</u>
ru ¹ , h ¹ , th ¹ , st ¹ , cu ¹ , ea ¹	35% (59)	25% (44)
TM8, Sb, e	40% (70)	0
	<u>P1</u>	<u>TM6C, Sb</u>
ru ¹ , st ¹ , ea ¹⁴ , spz ³ , ca ¹	29% (60)	39% (81)
TM1	32% (66)	0

Table 4.4. easter complementation analysis. *Drosophila* crosses were performed at 25°C and emerging adults scored from 10 days onwards until all had emerged. Percentages are the percentage number of flies of each genotype resulting from the cross. Numbers in parentheses are actual number of adult flies.

4.17. Conclusions from the P1 line analysis

Based on the analysis and discussion in **section 4.27.4** (see below), the overall conclusion is that the CG6156 gene has been disrupted in the P1 line and will be very unlikely to produce functional protein, but that additional genes are also likely to have been disrupted/deleted and not be producing protein. The homozygous lethality associated with this line cannot, therefore, necessarily be attributed to MCC disruption.

4.18. Broadening of the PCR screening

Since the P-element hop that had produced P1 had not generated a clean MCC mutant, a new strategy was employed. The hop lines were once again screened by PCR, but this time to search for insertions close to, but not within, the MCC gene itself. It was hoped that insertions would be identified that could be used for imprecise P-element excision in order to remove the MCC gene and a small section of intervening DNA, or to provide a position closer to MCC as the starting point for a second hop. Screening was carried out using the same genomic DNA preps as previously. New primers were designed to cover a ~2Kb region on each side of CG6156 (**fig. 4.16**) and these were used in PCRs with the primer at each end of the P-element, as before.

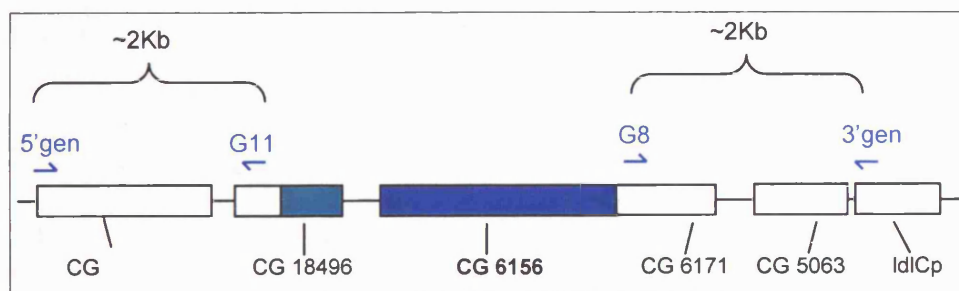


Figure 4.16. Broad screening genomic PCR primers. Not to scale

This round of screening identified one further insertion. A PCR product of approximately 2.2Kb was observed using primers G8 and PlacW3' (**fig. 4.17**). This was narrowed down to one particular line, and the insertion was named P2.

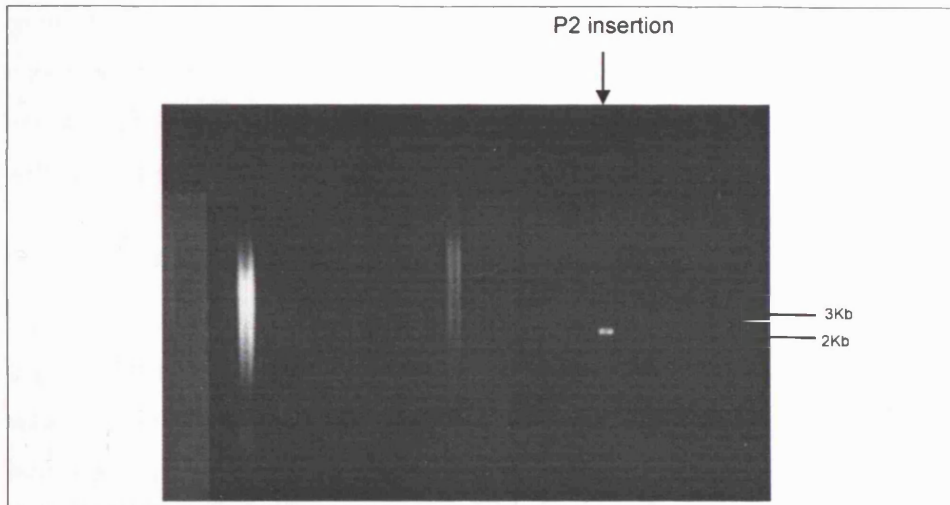


Figure 4.17. P2 insertion discovery. Genomic DNA from pools of 20 P-element hop flies was extracted and subjected to PCR using primers that covered the region either side of the MCC gene. Screening gel for lines 1-80.

This insertion is predicted to be within the *IdlCp* gene based on the size of the PCR product (fig. 4.18). Flies carrying the P2 insertion are homozygous viable with the same homozygous phenotype as the P0 insertion.

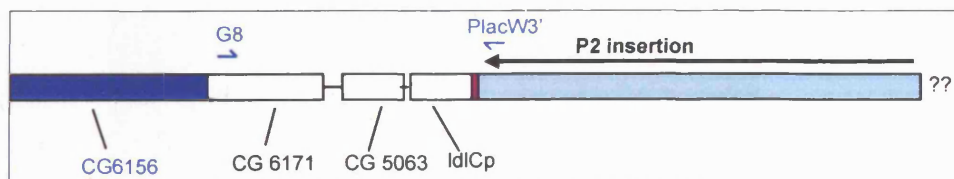


Figure 4.18. The P2 insertion.

4.19. The P2 insertion may be incomplete

Similarly to the P1 insertion, the P2 insertion appeared to be incomplete and/or there had been a genomic deletion. No PCR product was generated using the primer at the 5' end of the P-element and any of a number of genomic primers that should have given PCR product if an intact P-element were present. A plasmid rescue was performed, but once again all sequence corresponded to the original insertion, indicating that this is still present in the P2 line. This indicated that any deletion of the P-element in P2 was likely to have removed the ampicillin resistance

gene, and also indicated that there were likely to be no further intact insertions in this line except the P0 insertion. The eye colour of the P2 line was the same as the P0 line, indicating that the mini-white gene within the P2 P-element was unlikely to be intact.

4.20. The second hop

Despite the fact that the second insertion may be incomplete (and therefore the P2 P-element may not be competent to hop), a second hop was carried out using the P2 insertion line in an attempt to get a clean insertion into the MCC gene. It was predicted that even if the P2 insertion itself could not hop, the P0 insertion that is still present in this line could hop and therefore more insertions would be generated. Once again, approximately 1500 fly lines were screened for an insertion into the MCC gene (the 2Kb either side of the gene was not re-screened), and one further insertion was identified by PCR (**fig. 4.19**). This insertion (named P3) gave a PCR product of 500-700bp with the G9 + LacW5' primers.

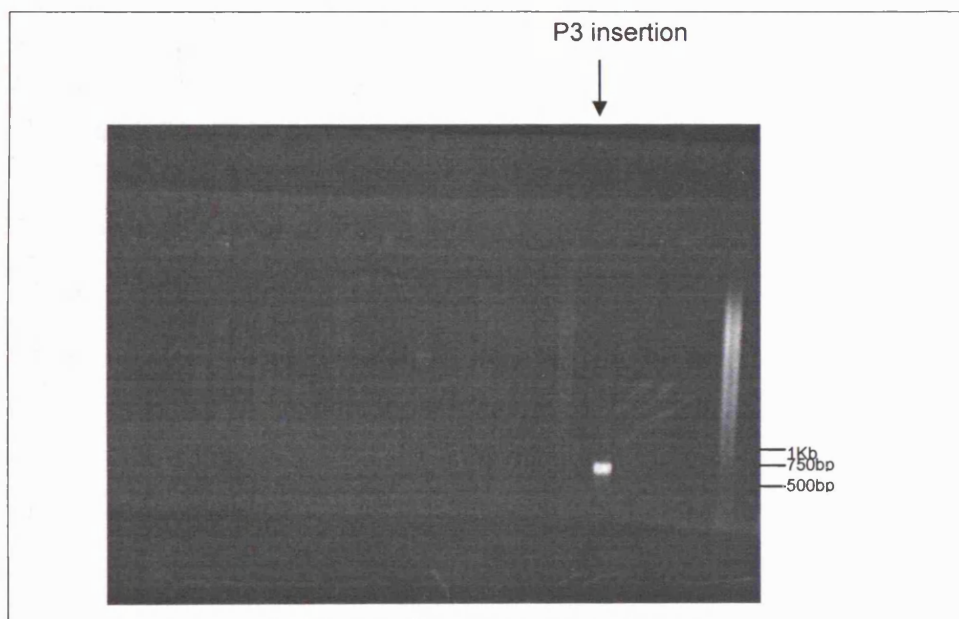


Figure 4.19. Discovery of the P3 insertions. Genomic DNA from pools of 20 P-element hop flies was extracted and subjected to PCR using primers that covered the MCC gene. Screening gel for lines 181-300.

4.21. Two distinct lines

As with the two previous insertions, the precise line carrying the insertion was identified by genomic DNA extraction of the 20 possible lines and

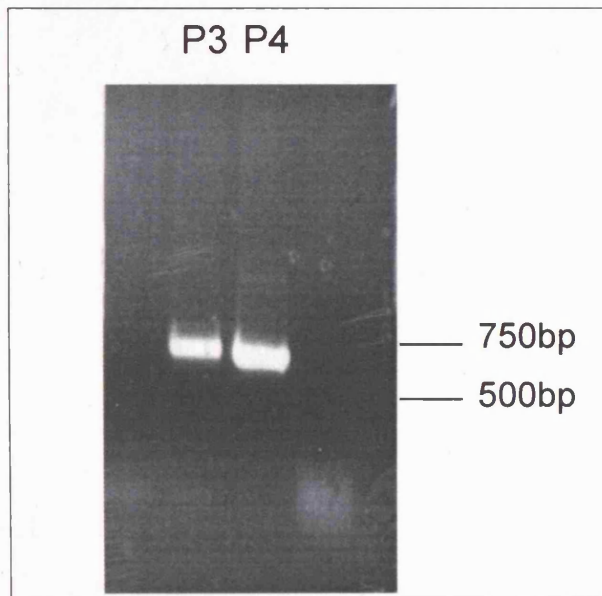


Figure 4.20. Isolation of the P3 and P4 insertion lines. Genomic DNA from all 20 P-element hop flies was extracted and subjected to PCR using primers G9 and LacW5'.

individual PCRs. In the case of the P3 insertion it was determined that two lines within the cohort of 20 gave PCR product with the G9 + LacW5' primers (fig. 4.20). The size of the PCR products for the two lines is subtly different indicating that they are likely to be two separate lines rather

than a duplication of the same line. These insertions are therefore called P3 and P4. Both P3 and P4 lines are

homozygous viable with the same bristle phenotype as the homozygous P0 line in each case (fig. 4.21).

It was also determined by PCR that the P2 insertion is still present in the P3/4 lines, and that the P0 insertion is still present in the P2 line and the P3 line (fig. 4.22). The PCR to test for the P0 insertion in the P4 line did not give a product despite the P4 line having the same homozygous phenotype as the P0 original insertion line.

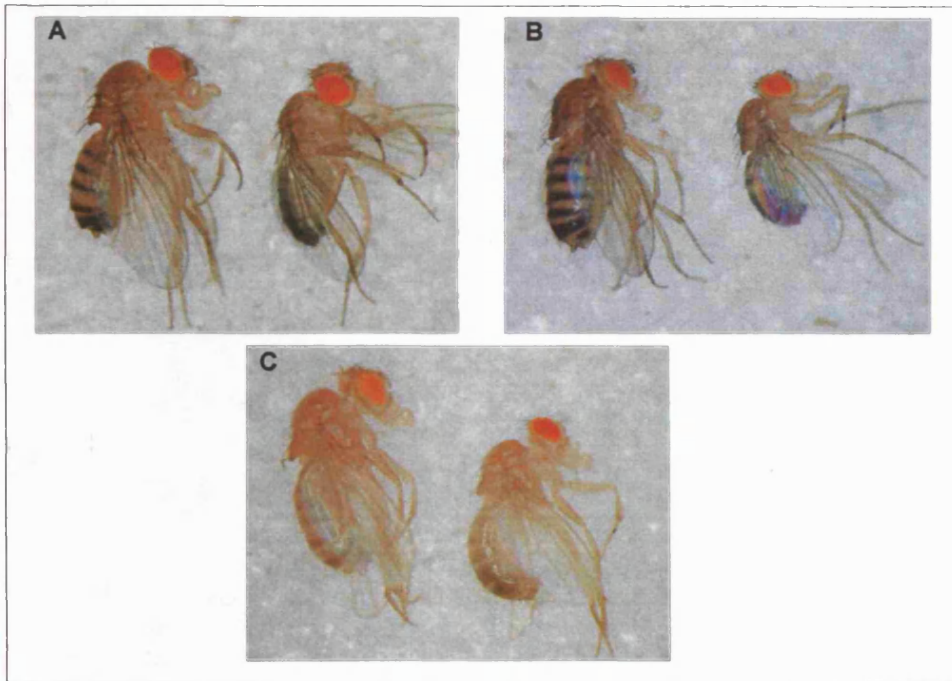


Figure 4.21. The P3 and P4 homozygous phenotype. A: P3 line, B: P4 line, C: P0 line. All images are of homozygotes.

4.22. The P3/4 P-element may be incomplete

Once again, the P-element insertion in MCC gave PCR product with primers at one end of the insertion (5' end), but not the other. As for the P1 insertion, the extent of the remaining P-element was mapped using a genomic primer (G9) in combination with the LacW5' primer. Once again, at least 8.8Kb of the P-element is still present at the insertion site, but the extreme 3' end of the P-element appears to have been deleted (E. Batchelor, personal communication). **Figure 4.23** is a map of the P-element insertions generated by the P-element hops as predicted by the PCR results.

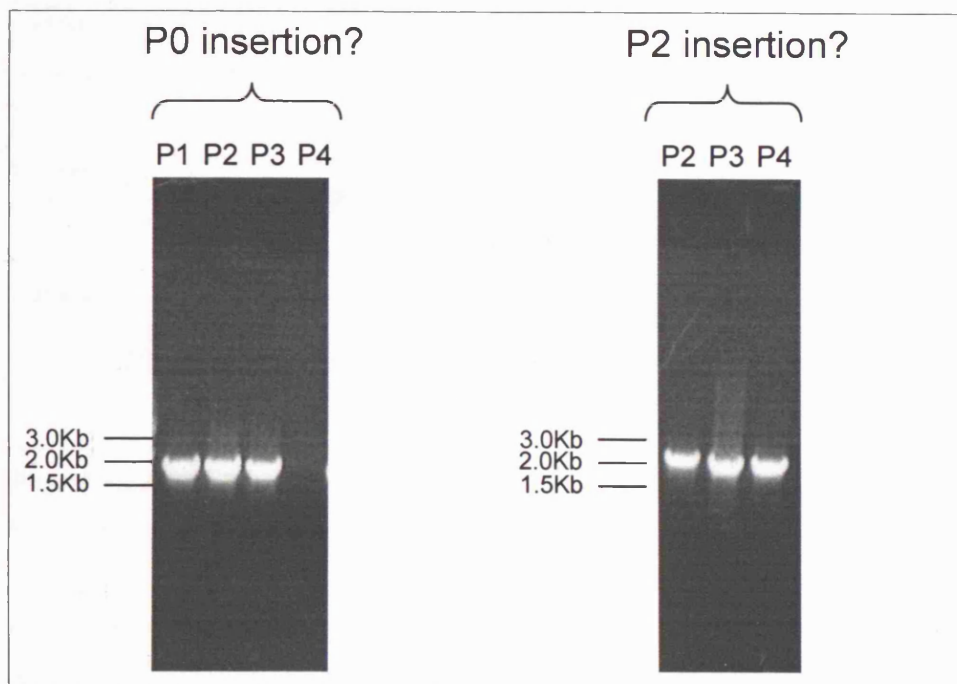


Figure 4.22. Testing for the presence of the original P-element insertions.

Genomic DNA from the individual lines was tested for the P0 insertion using primers PlacW5' + Taq3, and for the P2 insertion using primers G8 + PlacW3'.

4.23. Verifying the insertions

PCR can amplify DNA in a non-specific manner and generate a product that is not desired. In order to confirm that insertions are indeed present in each of the lines at the positions indicated in **figure 4.23** a Southern blot analysis was carried out. Southern blotting is a method to identify particular DNA sequences of interest within a large number of DNA fragments such as restriction digestion products of genomic DNA. Here, two probes were used for the Southern blotting (**fig. 4.24**).

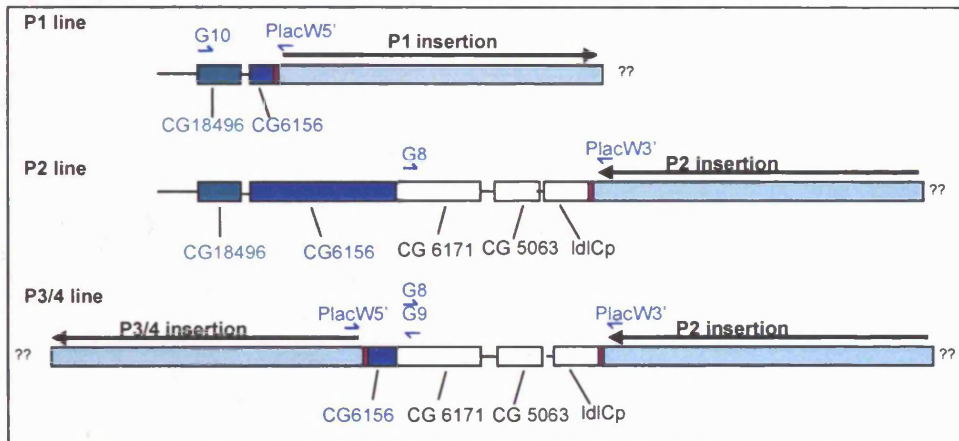


Figure 4.23. The P-element insertions as predicted by PCR analysis. The P0 insertion is not shown.

If a P-element has inserted into the region covered by the probe, different and/or additional bands to the wild-type bands should be visualised on the Southern blot. Two different restriction enzymes were used to generate fragments for the Southern analysis in the hope that bands distinct from the balancer chromosome wild-type bands would be clearly visible with at least one enzyme. The band sizes predicted by identifying restriction sites in the BDGP genomic sequence are depicted in **figure 4.25**.

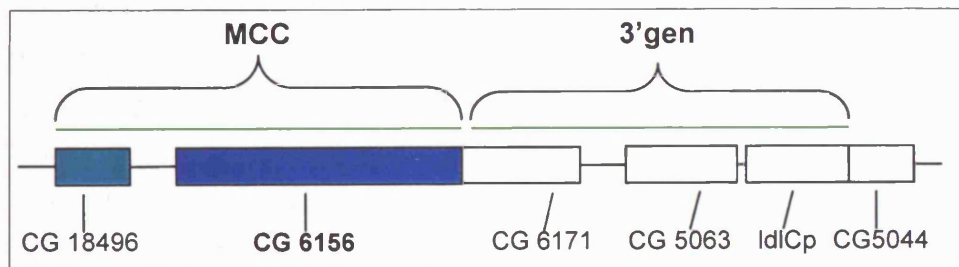


Figure 4.24. Probes for the Southern blot analysis. DIG-labelled probes were generated using standard methods (see Materials and Methods). The MCC probe covers the whole of CG6156 and CG18496, and was generated using primers G9 + G10. The 3'gen probe used primers G8 + gen4Kb and covers a region of ~4Kb directly to the right of CG6156. Not to scale.

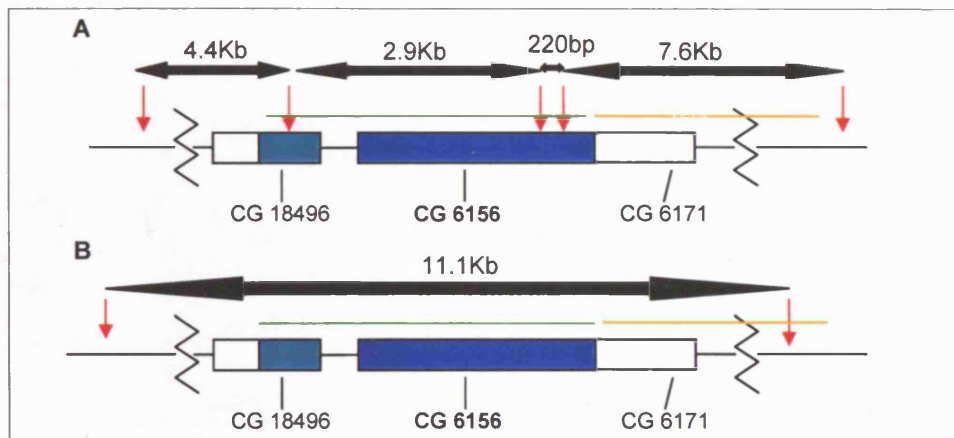


Figure 4.25. Prediction of approximate wild-type band sizes for the Southern blot analysis. A: BamHI digestion, B: HindIII digestion.

Green line indicates MCC probe, orange line indicates 3'gen probe, red arrows indicate restriction enzyme cut sites. Not to scale.

The Southern blot clearly indicates extra bands in the P1 and P4 insertion lines, confirming the presence of a P-element within the MCC gene locus in each case (fig. 4.26). The P3 genomic DNA digested with HindIII was unfortunately degraded (lane 4). However, the P3 insertion does not show any extra bands with BamHI (lane 4) (nor with ECoRI – not shown). This indicates that the P3 insertion line almost certainly does not carry a P-element within MCC.

In order to analyse a potential deletion of genomic DNA to the right of the 3' end of CG6156 in the P1 and P2 lines, the Southern blot was stripped and re-probed with the 3'gen probe (fig. 4.27). Blotting with this probe revealed no extra bands in addition to the predicted wild-type bands for the P2 line. This probably indicates that the P2 line does not carry a P-element at the predicted position. However, it is possible that an insertion at this point happens to give a new band(s) of the same size as the wild-type bands and is therefore not distinguishable. Analyses of the bands produced by the P3 or P4 lines (which are predicted by PCR to carry the P2 insertion, fig. 4.22) confirm that there is no insertion at the P2 predicted insertion site.

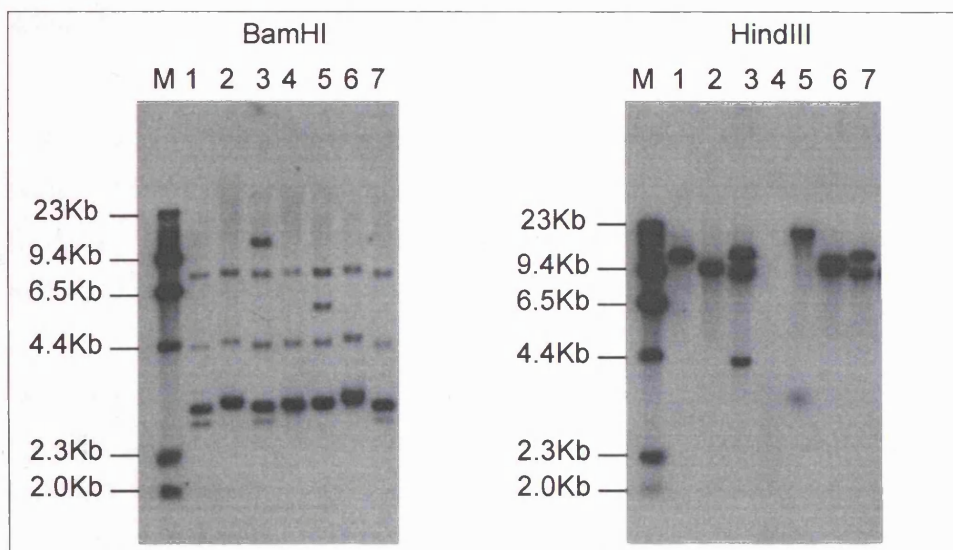


Figure 4.26. Southern blot of P-element insertions using the MCC probe. Total genomic DNA was extracted from the P-element insertion lines, digested with either BamHI or HindIII and subjected to Southern blot analysis using the MCC probe. P1 heterozygous flies and P2, P3 and P4 homozygous flies were used. OreR and P0 homozygous flies were used as controls. **M**: DIG-labelled markers, **1**: OreR, **2**: P0/P0, **3**: P1/TM3, **4**: P3/P3 **5**: P4/P4 **6**: P2/P2 **7**: Df(3R)ea/TM3.

A detailed analysis of the bands on the Southern blots and conclusions resulting from the analysis is provided in sections 4.27.2 -4.27.5 at the end of this chapter. In summary, the P4 insertion appears to be a clean insertion into MCC.

4.24. The P1 line is viable over the P4 line

Crossing P1 with P4 produced viable flies with no visible phenotype other than the bristle phenotype associated with the P0 insertion (table 4.5, fig. 4.28). Similarly, control crosses of P2 with either P1 or P4 produced the P0 homozygous phenotype, as did crosses of P1, P2 or P4 with P0 (data not shown).

4.25. The P1 and P4 lines complement deficiencies that take out MCC

The P1 and P4 lines were tested for complementation with two deficiency lines. These lines are both predicted to have removed the CG6156 and CG18496 genes, and all the genes from the 3' end of CG6156 up to and including *easter* (all these genes lie within 88F1):

Df(3R)ea (BL 383) – predicted breakpoints: 88E7-13;89A1

Df(3R)Exel6174 (BL 7653) – molecularly predicted to delete all genes from 88F1-88F7

Both lines are viable over both of these deficiencies with no visible phenotype (table 4.5, fig. 4.28).

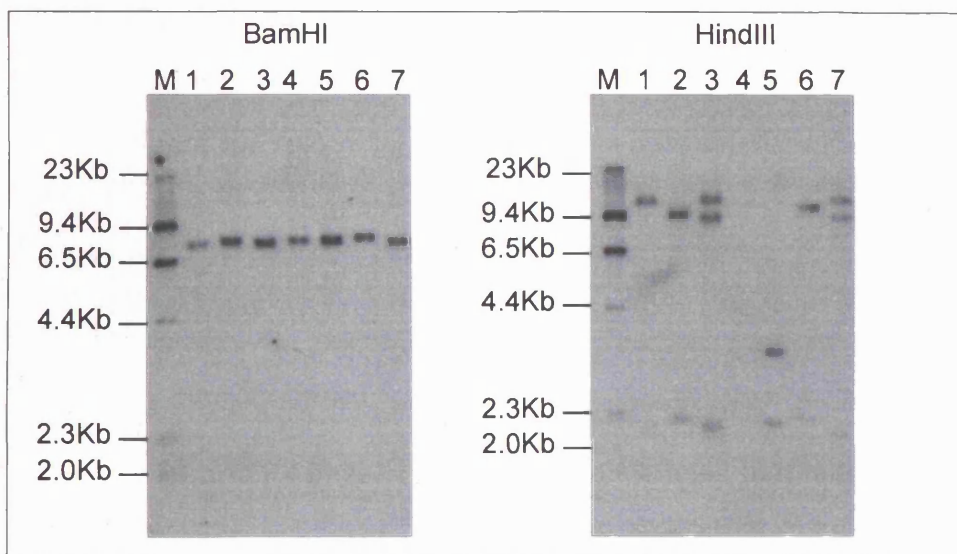


Figure 4.27. Southern blot of P-element insertions using the 3'gen probe.

Total genomic DNA was extracted from the P-element insertion lines, digested with either BamHI or HindIII and subjected to Southern blot analysis using the 3'gen probe. **M**: DIG-labelled markers, **1**: OreR, **2**: P0/P0, **3**: P1/TM3, **4**: P3/P3, **5**: P4/P4, **6**: P2/P2, **7**: Df(3R)ea/TM3.

4.26. Fertility defects

Although there was no phenotype observed in homozygous adults of the P4 line, it is possible that this insertion could lead to a defect in oogenesis, spermatogenesis, or some other process associated with fertility. If this were the case, phenotypes would only become visible in the second homozygous generation. In order to test this, an attempt was made to keep the P4 line as a homozygous stock.

	P4	TM6C, Sb
P1	(118) 31%	(134) 36%
TM6C, Sb	(124) 33%	0
	Df(3R)ea	TM3, Ser
P1	(97) 49%	(101) 51%
TM3, Sb, Kr-GFP	0	0
	Df(3R)Exel6174	TM6B, Hu
P1	(77) 50%	(78) 50%
TM6C, Sb	0	0
	Df(3R)ea	TM3, Ser
P4	(101) 39%	(98) 38%
TM6C, Sb	0	(58) 23%
	Df(3R)Exel6174	TM6B, Hu
P4	(108) 49%	(112) 51%
TM6C, Sb	0	0

Table 4.5. P-element complementation analysis.

Drosophila crosses were performed at 25°C and emerging adults scored from 10 days onwards until all had emerged. Percentages are the percentage number of flies of each genotype resulting from the cross. Numbers in parentheses are actual number of adult flies.

The P2 line was used as a control with the same genetic background as P4 but lacking a P-element insertion in MCC. The P0 line can be

maintained as a homozygous stock and was also used as a control. Neither P2 nor P4 flies could be maintained as a homozygous stock. This may imply that the P2 line carries a P-element insertion somewhere in the genome that is affecting fertility since P2/P2 homozygous flies are viable, but their progeny are not. This is presumably also present in the P4 line. P2 homozygous females do lay eggs but no larvae hatch from these eggs implying that they are either not fertilised, or are not competent to be fertilised.

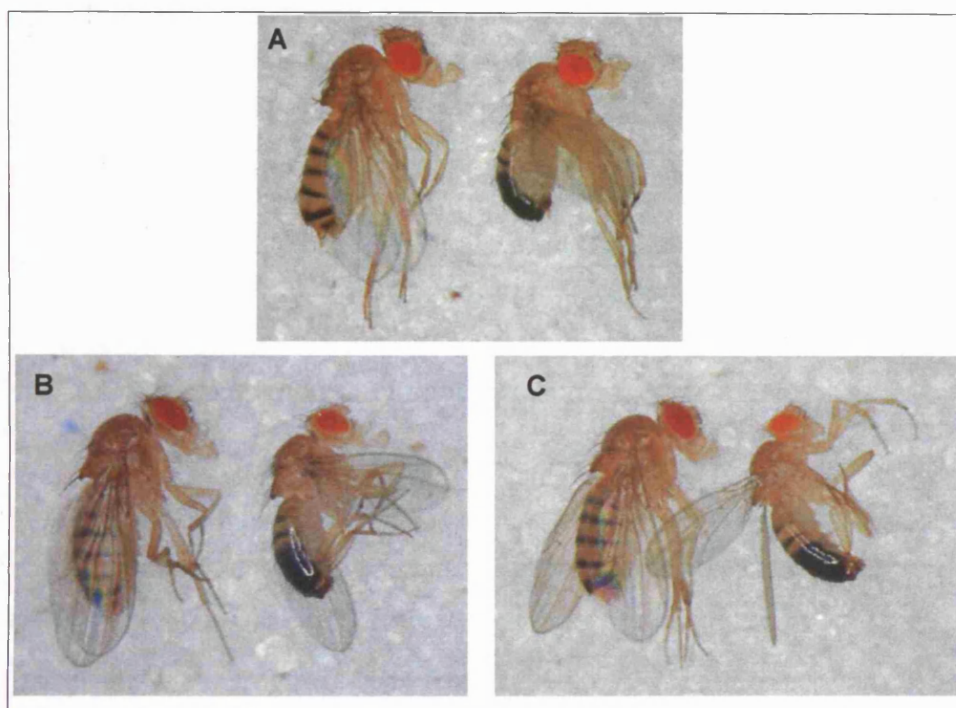


Figure 4.28. Complementation analysis of P-element lines. For details of methods see **table 4.5** legend. **A:** w/+ ; P1/P4, **B:** w/+ ; P1/Df(3R)ea, **C:** w/+ ; P4/Df(3R)ea.

The P4 and P2 lines were also crossed to the P1 line and attempts made to generate progeny from interbreeding P1/P4 or P2/P4. **Table 4.5** illustrates the results of these crosses. The lack of progeny from all of these crosses indicates that any fertility defect is not associated with an insertion in MCC, but is likely to be a separate mutation on the chromosome that is present in the P1, P2 and P4 lines but not the P0 line.

Line	Larvae?
P0/P0	YES
P4/P4	X
P2/P2	X
P1/P4	X
P1/P2	X
P2/P4	X

Table 4.6. Fertility crosses for P-element insertions. Crosses were carried out at 25°C. Each line was inter-bred with itself and the food observed for churning (associated with larval feeding) and adult emergence. X indicates churned medium was not observed.

4.27. Discussion and conclusions

4.27.1. RNAi and UAS-MCC expression studies

The lack of phenotype observed with the UAS over-expression crosses could indicate that there is a technical problem with the construct or that MCC over-expression has no effect on development. It is possible that the fly lines had lost the UAS insertions. They were tested by PCR with a primer internal to the hsp70 gene in combination with one internal to the MCC gene, and one internal to the white gene in combination with one internal to the MCC gene (both the hsp70 and white genes are present in PUASt, the vector used to make these transgenic lines), and all lines tested positive for the insertion. Another possibility is that the expression level achieved is not sufficiently high to give a phenotype. To circumvent this possibility, the crosses were also carried out at 29°C to achieve high GAL4 expression levels and therefore high MCC expression levels. This still produced no visible phenotypes.

There can be insertion position effects on expression levels of P-element mediated injected constructs (Spradling and Rubin 1983). It is possible that expression levels at some UAS-MCC insertion loci are very low, and therefore multiple insertions are required to provide sufficient over-expression levels to produce a phenotype. All lines have at least one insertion on the third chromosome, but it is not known whether multiple insertions are present. Lastly, it may be that the UAS construct is simply not functioning (i.e. it is not expressing MCC in response to GAL4 binding to the UAS). It would be possible to test this by carrying out an *in situ* of embryos carrying a GAL4 construct and UAS-MCC construct that would lead to ectopic MCC expression that could be assayed by use of the MCC probe. For example, using a wingless-GAL4 driver, a striped *in situ* pattern would be indicative of the UAS construct functioning correctly.

The final possibility is that over-expression of MCC produces no phenotype. If MCC is involved in blocking the cell cycle in its role as a tumour suppressor it is possible that its over-expression would have no cellular effect, and an effect would only be seen when the gene function was taken away. However, in this regard one would expect there to be a phenotype when using RNAi or in a mutant.

The RNAi construct covered exons E and F and should, therefore, lead to degradation of all four mRNAs since they all include at least part of these exons. The lack of phenotype could be due to similar technical reasons as for the UAS over-expression construct. In this regard crosses were also performed at 29°C. Fly lines with dark red eyes (potentially indicative of multiple mini white genes and therefore multiple insertions), and one with insertions on multiple chromosomes were chosen for the RNAi analysis in an attempt to select lines achieving high expression levels of the construct. RNAi constructs are often prone to difficulties with achieving high protein knock-down levels. Without an antibody it is not possible to ascertain the efficacy of the flipback RNAi construct used

here. However, it may be possible to visualise a knock down of mRNA levels either by *in situ*, Northern analysis or RT PCR.

4.27.2. P-element hopping (general comments)

Deletion of the 3' end of the P-element appeared to occur with both the P1 and P4 insertions. It is possible that the P-element in the P0 line that was the starting point for the hops is defective in some way that leads to a deletion of its extreme 3' end when it hops. Deletions are commonly associated with P-element hops and may solely involve regions internal to the P-element or in addition remove genomic DNA in the region of insertion (Staveley, Heslip et al. 1995).

4.27.3.Southern blotting (general comments)

The Southern blot revealed an interesting phenomenon when analysing the wild-type bands for OreR (lane 1, **fig. 4.26**). There appeared to be an extra band of approximately 2.7Kb in the BamHI digestion. This band could represent a polymorphism of the chromosome in the region of MCC. A polymorphism is a sequence difference observed between different chromosomes in the wild-type population. A single organism may have one chromosome with one version of the polymorphism and the homologous chromosome with a different version. If the polymorphism happens to remove or insert a restriction site this will be seen with Southern blotting using the corresponding restriction enzyme. The polymorphism is most obvious when a 5'gen probe is used (**fig. 4.29**). The polymorphism is also seen in the TM3 balancer chromosome (lanes 3 and 7 of **fig 4.29**). In this case an extra band is also observed with Hind III, indicating that there is also a polymorphism affecting a Hind III site (**fig. 4.26**).

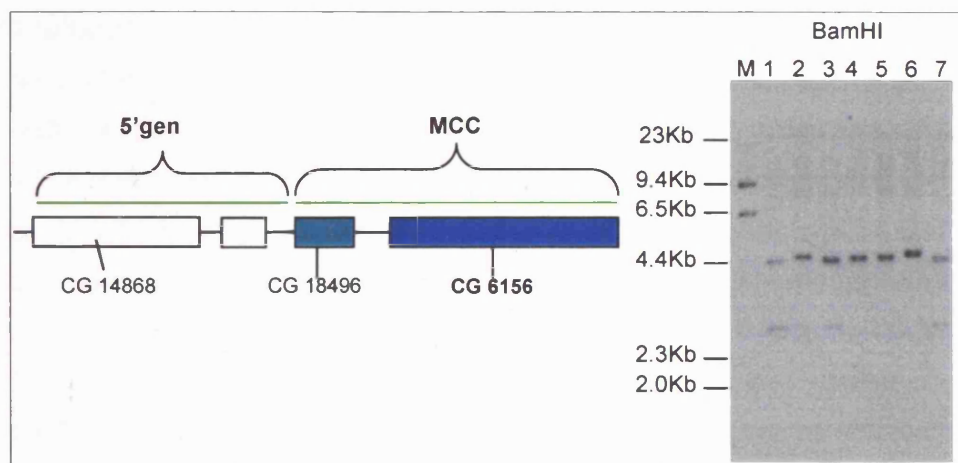


Figure 4.29. Polymorphism to the left of the MCC gene. Total genomic DNA was extracted from the P-element insertion lines, digested with either BamHI or HindIII and subjected to Southern blot analysis using the 5'gen probe. **M:** DIG-labelled markers, **1:** OreR, **2:** P0 line, **3:** P1/TM3, **4:** P3/P3, **5:** P4/P4, **6:** P2/P2, **7:** Df(3R)ea/TM3.

4.27.4. P1 insertion analysis

The P1 line is homozygous lethal. The insertion is located within the third exon of the CG6156 predicted gene. This exon is upstream of the end of mRNA_d, therefore no stable MCC mRNAs are likely to be produced from the P1 insertion chromosome.

The Southern analysis confirms the presence of a P-element in the region indicated by the PCR product sequencing. The PCRs indicate that the 3' end of the P-element and/or the 3' end of CG6156 (and genes to the right) may be deleted, but that at least 8.8 Kb and less than 10.5Kb of the P-element is present. The inverse PCRs designed to enable sequencing across the breakpoint were not successful possibly due to inefficient ligation. They should have given a PCR product from the P0 insertion that is present in the P1 line but they did not, indicating that the technique was not working.

The Southern analysis confirms that the 3' end of CG6156 has been deleted since there are no extra bands corresponding to this part of the MCC probe with either BamHI or HindIII unless they all happen to be the same size as the wild-type bands (which is unlikely). The BamHI bands suggest that a large proportion of the P-element is present in the P1 line since the extra band is large at ~13Kb (fig. 4.26). The genome contains many BamHI sites (on average 1 every 3.8Kb in the MCC region), but the P-element contains only one BamHI site almost at the 3' end. Based on the PCR analysis, and the fact that the plasmid rescue gave no colonies corresponding to the P1 insertion, it is likely that the P-element is incomplete and that the breakpoint is 3' of 8.8Kb and 5' of 10.5Kb within the P-element. The BamHI site giving rise to the ~13Kb fragment is, therefore, likely to be genomic and the fragment will include ~8.8Kb of the P-element (fig. 4.30).

Unfortunately, due to P1 being homozygous lethal, the Southern analysis tells us little about a potential deletion of genomic DNA to the right of CG6156 since the wild-type bands are always present. The only information we have about this is the genetic analysis with the *easter* mutants that indicate that the *easter* gene is intact within this line. Therefore, the conclusion from analysis of the P1 line is that the P-element has inserted within the third exon of CG6156 and that the part of CG6156 downstream from the insertion point has been deleted from the genome along with a section of the chromosome that does not reach as far as the *easter* gene.

The homozygous lethality of P1 does not fit with its viability over both deficiencies that take out all genes from CG6156 to *easter*. If deletion of the 3' end of CG6156 itself or any of the other genes removed in the P1 line were to give rise to a homozygous lethal phenotype, it would be expected that the combination of P1/deficiency would be homozygous lethal too. There are several possible reasons for this discrepancy. One explanation is that the deficiencies have not been accurately mapped. Df(3R)ea breakpoints have been mapped cytologically, a technique with a

high degree of error. Df(3R)Exel6174 has been mapped molecularly which should ensure accuracy of claimed deleted genes. However, re-mapping of the deficiency would be required to confirm this data. The deletion in the deficiency could be confirmed by either a Southern blot or PCR approach. If either of these deficiencies do not remove CG6156 and/or any other genes that have been deleted in the P1 line, they could be viable in combination with P1.

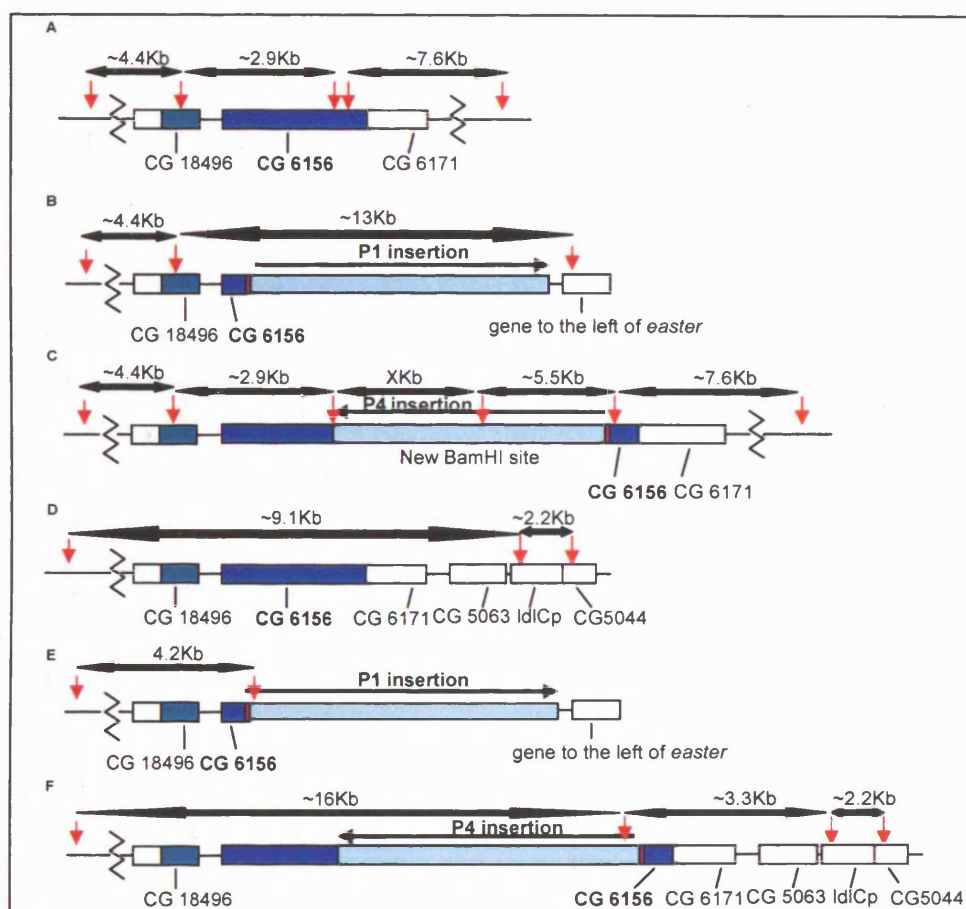


Figure 4.30. Hypothesis for P1 and P4 insertion outcomes and resulting Southern bands. A-C: BamHI digestion, **D-F:** HindIII digestion. **A, D:** wild-type chromosome (e.g. balancer), **B, E:** P1 insertion chromosome, **C, F:** P4 insertion chromosome. Red arrows indicate restriction enzyme cut sites. Not to scale.

A second possible reason why P1 is homozygous lethal but viable over the deficiencies is that there is another P-element insertion somewhere in the genome that leads to homozygous lethality. If this were the case it would presumably have a non-functional ampicillin resistance gene (as the P1 insertion does) since the plasmid rescue only gave colonies with sequence corresponding to the original insertion. It is possible that the combination of P0 and P1 insertions gives rise to lethality, which would explain why the P1 line, which carries both insertions, is homozygous lethal but viable over the deficiencies. However, both deficiencies are predicted to remove the insertion point of P0, so it seems unlikely that it is the combination of P0 and P1 that gives rise to lethality. It is more likely to be a third insertion elsewhere in the genome.

The overall conclusion of P1 line analysis is that the CG6156 gene has been disrupted and will be very unlikely to produce functional protein, and that additional genes are likely to have been disrupted/deleted and not be producing protein. The homozygous lethality associated with this line cannot necessarily be attributed to MCC disruption.

4.27.5. P4 insertion analysis leading to the conclusion that an MCC mutant has no obvious phenotype

The P4 line is homozygous viable with the same phenotype as the P0 line. The P4 insertion is predicted by the size of the PCR products to be located within exon H (see **fig. 4.10**) just to the right of the stop codon used in DMCCd. It is possible that mRNA_d is expressed in the P4 line. The RT PCR in chapter 3 did not define the 3' end of mRNA_d, but it is likely that it continues after the stop codon with a 3' UTR. The likelihood is, therefore, that the P4 insertion line is not able to produce stable mRNAs since the coding region would be separated from the 3' untranslated region and site of polyA tail addition, which are needed to confer stability (Wickens, Anderson et al. 1997). The P4 insertion is viable over both deficiencies (although, as discussed above, these may not remove CG6156).

PCRs indicate that the 3' end of the P-element and/or the 5' end of CG6156 (and upstream genes) may be deleted but that at least 8.8 Kb of the P-element is present in the P4 line. The Southern analysis confirms the predicted region of P-element insertion. Confusingly, the BamHI analysis of the P4 line apparently shows wild-type bands in addition to the non-wild type band of approximately 5.5Kb. A likely explanation for this is that the P-element has inserted between the two closely spaced BamHI sites towards the 3' end of the MCC gene (fig. 4.30). These are located only 266 bp apart and from the size of the PCR product using the G9 + PlacW5' primers (between 500 and 750bp) it is possible that the P-element has inserted between them. This would give rise to wild-type size bands in addition to the extra band. Sequencing of the PCR product would confirm this.

The Southern analysis of P4 using BamHI indicates that less than 5.5Kb of the P-element is present since P(lacW) contains only one BamHI site almost at its 3' end and a 5.5Kb band is observed. However, digestion with HindIII gives a very large band of approximately 16Kb which is larger than a combination of both the P-element and the whole CG6156 gene (the P-element contains HindIII sites at both its 5' and 3' ends). These two facts are contradictory, and there must be something more complicated going on. One hypothesis would be that the P-element has gained a BamHI site approximately 5.5Kb from its 3' end giving rise to the 5.5Kb BamHI site (we know that the P-element in the P1 insertion does not contain a BamHI site at this position) (fig. 4.30). P-elements move via a cut-and-paste mechanism which is prone to errors and internal deletions are common (Staveley, Heslip et al. 1995). P-elements also replicate themselves, another error-prone process that may result in sequence alternations. Either an internal deletion or replication error may have given rise to a novel BamHI site. The fact that the 5.5Kb band is weak in intensity indicates that it is likely to correspond to the probe hybridising to the small region of CG6156 that is to the right of the P4 insertion. There is no other extra band corresponding to the 5' end of

CG6156/CG18496. This could be indicative of a deletion of the 5' end of MCC or of a BamHI site immediately to the left of the P4 insertion giving rise to a wild-type band size. If this BamHI site were sufficiently close to the end of the P-element there would only be a very small section of sequence that the probe would hybridise to, and therefore the band corresponding to that small section may be present on the gel but not visible with the probe. Depending on the extent of the 3' deletion of the P4 insertion (which is likely to be small based on the size of the HindIII fragment), it is possible that this band is the same size as one of the other bands on the blot (most likely the 5.5Kb or 2.9Kb bands) and is therefore not visible or that it is too faint to be seen for reasons just described.

The analysis of the P4 line using HindIII confirms that the 5' end of CG6156 is present since the ~16Kb fragment is dark, and that the P-element is likely to be mostly intact. Similarly to BamHI, the frequency of HindIII sites is high within the genome (in the MCC region they occur on average one every 3.25Kb), and therefore a large fragment is indicative of a mostly intact P-element insertion since the P-element only has two HindIII sites – one at each end.

In conclusion, in the P4 line a P-element missing only its extreme 3' end has inserted in the seventh exon of CG6156. This insertion is not likely to produce functional protein from the MCC locus, but no other genes should be directly affected (unless the P-element is within an enhancer of another gene). An MCC mutant, therefore, has no homozygous phenotype (the P4 line does, of course, have the background P0 phenotype). It is important to confirm that the P4 line does not express functional protein. Without an antibody against *Drosophila* MCC this is impossible.

4.27.6. Complementation analysis

The fact that P1 is homozygous lethal but is viable over the P4 insertion indicates that mutation of a gene other than MCC is giving rise to the

homozygous lethality since both P1 and P4 lines are unlikely to be producing functional MCC protein of any form. As discussed above, it is possible that an additional P-element insertion elsewhere in the genome is giving rise to P1 homozygous lethality.

There is a *Drosophila* line available that carries a P-element insertion within CG18496 (Bloomington stock 18571). This line is homozygous viable, and both the P1 and P4 insertions are viable over it (data not shown). This insertion would be predicted to disrupt expression of mRNA^c and mRNA^d. This is further evidence for lack of phenotype of an MCC mutant, although mRNA^a and mRNA^b may be expressed normally in this line.

4.27.7. Lack of phenotype

The lack of phenotype for an MCC mutant indicates that MCC protein has no obvious function during *Drosophila* development. The lack of phenotype seen with the mutant correlates with the lack of phenotype seen with RNAi analysis. In order to absolutely confirm lack of gene expression in the P-element lines, an antibody against MCC could be used to illustrate lack of MCC protein. It is just possible that an mRNA may be produced from either P1 or P4 lines that runs into the P-element and is stable. However, this is unlikely in itself, and would be even more unlikely to produce a stable and functional in-frame protein. Certainly in the case of the P1 insertion, any protein produced would contain only a short section of amino acids of MCC origin, and would be unlikely to be functional.

It is possible that the P-element mutant has a phenotype that is not obviously visible such as a behavioural phenotype associated with MCC CNS expression. This may not affect CNS development *per se* but affect its function once it is developed, and give rise to a subtle behavioural phenotype that could only be observed in specialised behavioural studies.

If MCC were acting as a tumour suppressor one might expect a visible phenotype, perhaps involving overgrowth when the gene is mutated, such as is seen with the leukaemia associated tumour suppressor (LATS) mutant (Xu, Wang et al. 1995). No overgrowth phenotype is observed, and therefore indicates that MCC is unlikely to be acting as a "tumour suppressor" in *Drosophila*. This does not mean, however, that MCC does not act as a tumour suppressor in higher organisms. The fact that MCC is expressed in tissues other than the central nervous system in higher organisms such as the mouse may indicate additional roles for MCC in those organisms.

If DMCC were functioning in the signalling pathway involving DRhoGEF2, a DMCC mutant would be expected to show a phenotype similar to the DRhoGEF2 mutant phenotype. Since this is not the case, it is unlikely that DMCC is functioning in the signalling pathway involving DRhoGEF2.

4.28. Chapter 4 References

- Bellen, H. J., R. W. Levis, et al. (2004). "The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes." Genetics **167**(2): 761-81.
- Bier, E., H. Vaessin, et al. (1989). "Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector." Genes Dev **3**(9): 1273-87.
- Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**(6669): 806-11.
- Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-51.
- Kalidas, S. and D. P. Smith (2002). "Novel genomic cDNA hybrids produce effective RNA interference in adult *Drosophila*." Neuron **33**(2): 177-84.
- O'Hare, K. and G. M. Rubin (1983). "Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome." Cell **34**(1): 25-35.

- Phelps, C. B. and A. H. Brand (1998). "Ectopic gene expression in *Drosophila* using GAL4 system." Methods **14**(4): 367-79.
- Rorth, P. (1998). "Gal4 in the *Drosophila* female germline." Mech Dev **78**(1-2): 113-8.
- Rubin, G. M. and A. C. Spradling (1982). "Genetic transformation of *Drosophila* with transposable element vectors." Science **218**(4570): 348-53.
- Spradling, A. C. and G. M. Rubin (1983). "The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene." Cell **34**(1): 47-57.
- Staveley, B. E., T. R. Heslip, et al. (1995). "Protected P-element termini suggest a role for inverted-repeat-binding protein in transposase-induced gap repair in *Drosophila melanogaster*." Genetics **139**(3): 1321-9.
- Tower, J., G. H. Karpen, et al. (1993). "Preferential transposition of *Drosophila* P elements to nearby chromosomal sites." Genetics **133**(2): 347-59.
- Wickens, M., P. Anderson, et al. (1997). "Life and death in the cytoplasm: messages from the 3' end." Curr Opin Genet Dev **7**(2): 220-32.
- Xu, T., W. Wang, et al. (1995). "Identifying tumor suppressors in genetic mosaics: the *Drosophila* lats gene encodes a putative protein kinase." Development **121**(4): 1053-63.

5. Results – DMCC and DRhoGEF2 interaction analysis

5.1. Introduction

Although it is unlikely that DMCC participates in the signalling pathway involving DRhoGEF2 leading to cell shape changes based on the expression and mutation analyses presented in chapters 3 and 4, DRhoGEF2 may have other functions for which DMCC is required. Alternatively, DRhoGEF2 could be influencing any function DMCC may have. In this chapter the interaction between DRhoGEF2 and DMCC is investigated in order to understand whether DMCC influences DRhoGEF2 function, or vice versa.

The yeast 2-hybrid result (**Introduction**) indicates a possible interaction between the DRhoGEF2 PDZ domain and DMCC. However, yeast 2-hybrid studies can give false positives, and confirmation of the interaction is required, which is presented here. The fact that the interaction is likely to be mediated via the DRhoGEF2 PDZ domain is interesting as it opens up questions of specificity. For example, does DMCC only interact with DRhoGEF2 PDZ domain or other PDZ domains too and what, molecularly, is required for the interaction? These issues are begun to be addressed here.

The functional interaction between DRhoGEF2 and DMCC is investigated by a genetic approach. Despite the fact that mutation of DMCC does not lead to a phenotype *per se* (chapter 4) it is possible that the DMCC mutant alleles could enhance or suppress a DRhoGEF2 phenotype indicating that a genetic interaction exists between DRhoGEF2 and DMCC.

5.2. Interaction between the isolated DRhoGEF2 PDZ domain and DMCC

In order to confirm the yeast 2-hybrid result, a co-immunoprecipitation strategy was used. The DRhoGEF2 PDZ domain and DMCC were

generated as fusion proteins with N-terminal tags. The recombinant PDZ domain co-immunoprecipitated DMCC from a *Drosophila* S2 cell lysate (fig. 5.1).

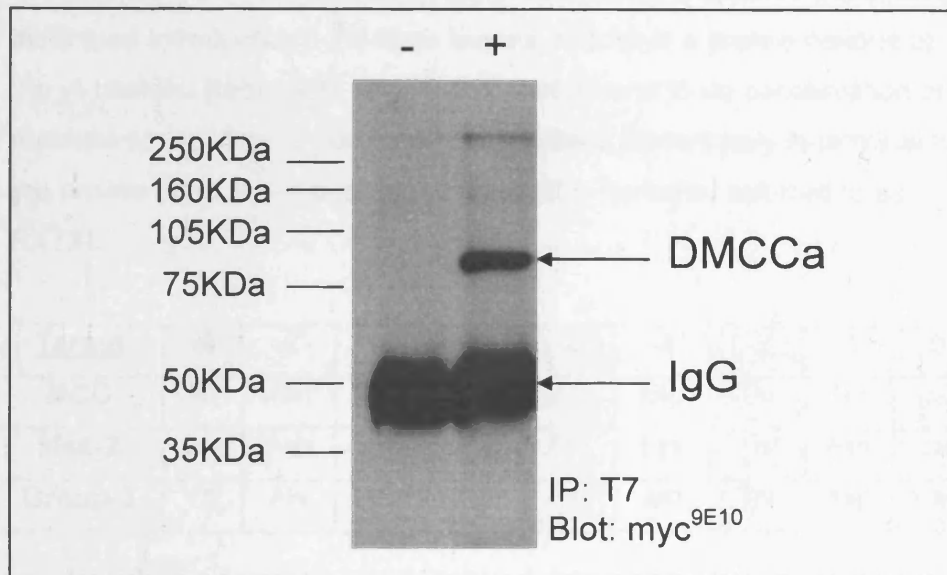


Figure 5.1. DMCCa co-immunoprecipitates with DRhoGEF2 PDZ domain.

The recombinant PDZ domain was generated using a bacterial recombinant expression system which provides a T7 tag at the N-terminus. Full-length DMCCa was cloned into a modified pFASTBAC1 vector (Invitrogen) which has the Act5C promoter in place of the Polh promoter, and provides a myc tag at the N-terminus. The Act5C promoter within pFASTBAC-Act5C gives high expression levels when transfected into *Drosophila* culture cells. pFASTBAC1-Act5C-myc-DMCCa was transfected into *Drosophila* Schneider S2 cells and cells were lysed prior to addition of the recombinant PDZ domain and immunoprecipitation with an anti-T7 antibody. The membrane was blotted with the anti-myc^{9E10} antibody. - no PDZ domain, + with PDZ domain.

The bright band at ~50KDa in **figure 5.1** is due to the secondary anti mouse-IgG antibody recognising the mouse T7 antibody used to immunoprecipitate the PDZ domain.

5.3. Conservation of PDZ-binding motif

The three potential DRhoGEF2 PDZ targets identified in the yeast 2-hybrid screen (MCC, group 3 and Mec-2) all have a Threonine – X – Leucine – COOH motif at their C-terminus. This is a classic PDZ-binding motif (see Introduction). All three targets also have a proline residue at the -4 position (table 5.1). There does not appear to be conservation of residues at the -1 or -3 positions, nor positions immediately N-terminal to the proline up to the -8 position. This motif is hereafter referred to as PXTXL.

<u>Target</u>	-8	-7	-6	-5	-4	-3	-2	-1	0
MCC	Gln	Asn	His	Val	Pro	Glu	Thr	Thr	Leu
Mec-2	Leu	Asp	Ala	Trp	Pro	Lys	Thr	Asn	Leu
Group-3	Val	Ala	Asp	Ala	Pro	Iso	Thr	Asp	Leu

Table 5.1. C-terminal amino acid residues of DRhoGEF2 predicted PDZ targets. 0 indicates the C-terminus.

5.4. Requirement of C-terminal residues for interaction

Since all DRhoGEF2 PDZ targets identified in the yeast 2-hybrid screen carry a PXTXL motif at their C-terminus, it is possible that the proline, threonine and leucine residues are necessary for binding to the DRhoGEF2 PDZ domain. In order to test the requirement of these residues for interaction with DRhoGEF2 PDZ domain, they were each individually mutated by site-directed mutagenesis of the the MCCa isoform to an alanine residue (K. Barrett). They were tested, as previously, for interaction with DRhoGEF2 PDZ domain using a co-immunoprecipitation approach (fig. 5.2).

Neither the DMCCa^{P→A} nor the DMCCa^{L→A} mutants co-immunoprecipitate with the DRhoGEF2 PDZ domain. The DMCCa^{T→A} mutant co-immunoprecipitates less effectively than wild-type DMCCa. DMCCd, which lacks the C-terminal PXTXL motif, and which was expressed at

very high levels (fig. 5.2, lane 6, blot C) also co-immunoprecipitates with DRhoGEF2 PDZ domain. This is likely to be due to excessively high expression levels and not represent a real interaction.

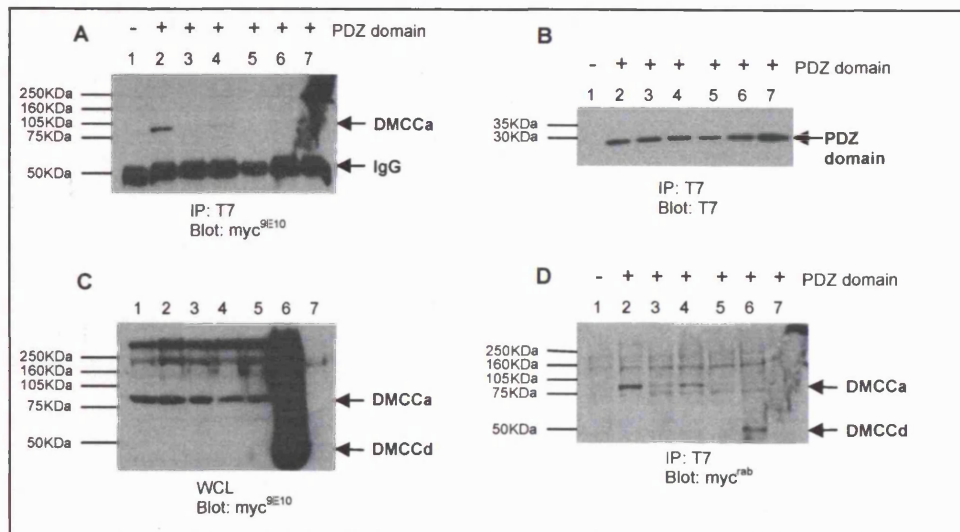


Figure 5.2. The proline and leucine residues are required for binding to DRhoGEF2 PDZ domain. DMCCa, DMCCa^{P→A}, DMCCa^{T→A} and DMCCa^{L→A} were individually cloned into pFASTBAC1-Act5C-myc and expressed in *Drosophila* S2 cells. Recombinant DRhoGEF2 PDZ domain with an N-terminal T7 tag was used for the IP. **A:** IP using anti-T7, blot using anti-myc^{9E10}. 1: DMCCa, 2: DMCCa, 3: DMCCa^{P→A}, 4: DMCCa^{T→A}, 5: DMCCa^{L→A}, 6: DMCCd, 7: untransfected control. **B:** IP using anti-T7, blot using anti-T7. Lanes as in A. **C:** Whole cell lysates (WCL), blot using anti-myc^{9E10}. 1: DMCCa, 2: DMCCa, 3: DMCCa^{P→A}, 4: DMCCa^{T→A}, 5: DMCCa^{L→A}, 6: DMCCd, 7: untransfected control. **D:** IP using T7, blot using a rabbit anti-myc antibody (anti-myc^{rab}). Lanes as in A.

5.5. Bioinformatics to identify further PXTXL proteins

If DRhoGEF2 does interact with targets based on sequence specificity for PXTXL at the C-terminus, it would be interesting to determine how many potential binding partners for DRhoGEF2 exist in *Drosophila*. A TXL-COOH motif occurs in 116 (PATSCAN) (appendix 3) or 123 (D.Sims, personal communication) proteins of the *Drosophila* proteome. This list is likely to include a large number that interact with PDZ domain proteins

since TXL is a classic C-terminal motif for a PDZ target protein. A PXTXL motif occurs in six of these (table 5.2).

<u>Name/ID</u>	-4	-3	-2	-1	0
MCC / CG6156	Pro	Glu	Thr	Thr	Leu
mec-2 / CG7635	Pro	Lys	Thr	Asn	Leu
group-3 / CG9795	Pro	Ile	Thr	Asp	Leu
capricious / CG11282	Pro	Val	Thr	Glu	Leu
CG5539	Pro	Gly	Thr	Ile	Leu
CG3777	Pro	Gln	Thr	Gly	Leu

Table 5.2. PXTXL-COOH proteins of the *Drosophila* proteome (BLAST analysis).

Therefore, if the DRhoGEF2 PDZ domain does select specifically for PXTXL in its binding partners, there are only six proteins within the cell with which it could potentially interact (see sections 5.9.2-5.9.4 for discussion of DRhoGEF2 PDZ domain specificity).

5.6. Interaction between full-length DRhoGEF2 and DMCC

In order to test the interaction between DRhoGEF2 and DMCC in a more physiologically relevant manner, the full-length proteins were over-expressed in *Drosophila* S2 cells and co-immunoprecipitation attempted following cell lysis (fig. 5.3).

Full-length DRhoGEF2 does not co-immunoprecipitate with DMCCc when the two full-length proteins are co-transfected in *Drosophila* S2 cells under the conditions tested (fig. 5.3). The addition of phosphatase inhibitors to push the system to an “active” phosphorylated state does not promote co-immunoprecipitation between DRhoGEF2 and DMCC, nor does treatment with GTPγS to lock G-proteins in an active state (data not shown), although co-transfection was not very efficient for these experiments (see section 5.9.1).

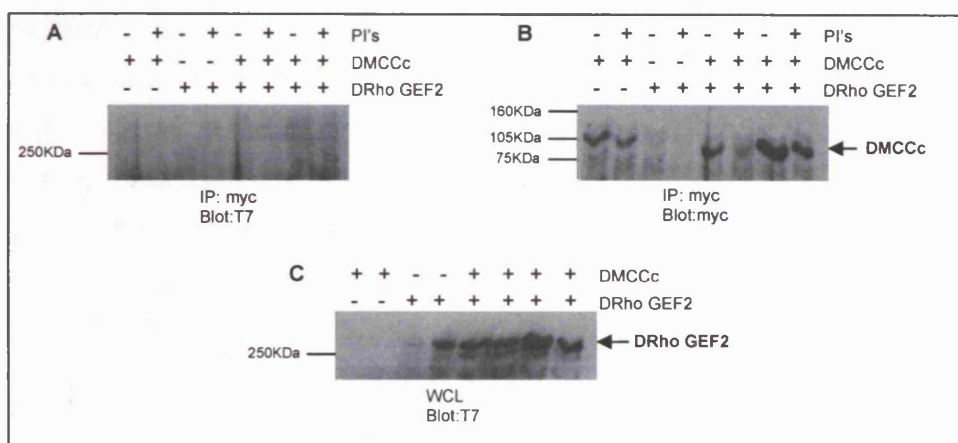


Figure 5.3. Full-length DRhoGEF2 does not co-immunoprecipitate with DMCC in S2 cell lysates. Three constructs were co-transfected into S2 cells: pAct5C-GAL4, pUASp-T7-DRhoGEF2 and pFASTBAC1-Act5C-myc-DMCCc as indicated. The pAct5C-GAL4 construct was used to express DRhoGEF2 from pUASp. The efficiency of co-expression was not high. In order to optimise this plasmids were co-transfected in different ratios and into different cell types (not shown). A 1:1:1 ratio transfected into *Drosophila* S2 cells was determined to be optimal and used here, although efficiency of co-transfection was still rather low. Anti-myc^{9E10} was used for immunoprecipitation and co-immunoprecipitation was tested by blotting with anti-T7 antibody. **A:** IP using anti-myc, blot using anti-T7, PI's – phosphatase inhibitor cocktail, **B:** IP using anti-myc, blot using anti-myc, PI's – phosphatase inhibitor cocktail, **C:** whole cell lysate blot using anti-T7.

There are various possible reasons why MCC was not shown to co-immunoprecipitate with full-length DRhoGEF2, and these are discussed in section 5.9.1. It is possible that the two proteins do interact *in vivo*, but that the conditions used in the experiments here are not conducive to binding. However, confirmation of MCC binding to full-length DRhoGEF2 would be required to illustrate that the interaction does happen in a cell.

5.7. Genetic interaction between DRhoGEF2 and DMCC

In order to test in another manner whether DMCC and DRhoGEF2 proteins are interacting *in vivo* a genetic interaction analysis was performed. Genetic interaction analysis involves the use of loss-of-

function or gain-of-function mutations in a specific gene that give a certain phenotype, and looks for modification of this phenotype by mutation in a distinct gene. In this case a phenotype caused by DRhoGEF2 loss of function was used in an assay for enhancement/suppression of phenotype by presumed loss-of-function mutations in MCC (the P-element insertions described in chapter 4). There were various DRhoGEF2 alleles used in the genetic interaction experiments (table 5.3).

Allele	Molecular abnormality	Comment
DRhoGEF2 ^{4.1} (4.1)	Breakpoint between amino acids 763 and 885 (between PDZ and C1 domains)	null
DRhoGEF2 ^{Px6} (Px6)	2132bp insertion in intron 1	hypomorph
DRhoGEF2 ^{Px10} (Px10)	860bp insertion in intron 1	weak hypomorph
DRhoGEF2 ^{6.5} (6.5)	Arg1687Cys	dominant negative

Table 5.3. DRhoGEF2 alleles. (Nikolaidou and Barrett 2004)

The combination of a DRhoGEF2 null allele (*DRhoGEF2*^{4.1}) and hypomorphic allele (*DRhoGEF2*^{Px6}) gives rise to a distinctive phenotype in *Drosophila*, most notably in the wings (fig. 5.4). The combination of a null allele (*DRhoGEF2*^{4.1}) and a much weaker hypomorph (*DRhoGEF2*^{Px10}) gives rise to a low penetrance phenotype of folded wings. These phenotypes are, of course, only seen in surviving flies - previous work has indicated that the viability to adulthood of 4.1/Px6 is 30% and 4.1/Px10 is 70% (Nikolaidou and Barrett 2004).

There were three different types of interaction experiment performed (for details of the crosses see section 2.16, Materials and methods):

1) Heterozygous DRhoGEF2 with heterozygous DMCC e.g.

w ; 4.1/+ ; P1/+ vs w ; 4.1/+ ; +/+

2) Homozygous DRhoGEF2 with heterozygous DMCC e.g.

w ; 4.1/Px6 ; P1/+ vs w ; 4.1/Px6 ; +/+

3) Homozygous DRhoGEF2 with homozygous DMCC e.g.

w ; 4.1/Px6 ; P1/P4 vs w ; 4.1/Px6 ; +/+

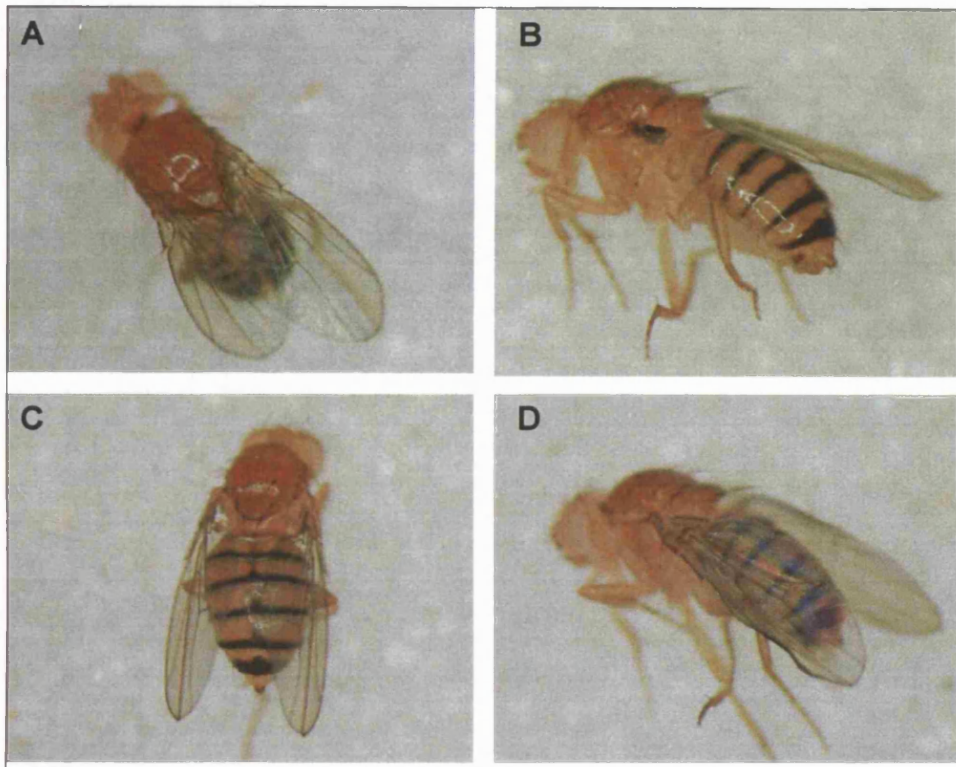


Figure 5.4. DRhoGEF2 null / hypomorph phenotypes. **A:** 4.1/Px6 - wing length difference is a weak phenotype for this genotype. **B:** 4.1/Px6 - wing stump is a stronger phenotype resulting from a folded wing getting stuck in the pupal case and being ripped off during adult eclosure. **C:** 4.1/Px10 - normal wings. **D:** 4.1/Px10 - slightly folded wing is a strong phenotype for this genotype.

Each experiment provides a progressively more sensitive background than the previous. It was hoped that DMCC mutation may enhance the phenotype of a weakly sensitive background but give rise to lethality in a

more sensitive background (for example a phenotype may be observed with experiment 1, and total lethality observed in experiment 2).

	4.1 ; +	CyO ; +
+ ; P1	26% (68)	19% (49)
+ ; +	28% (72)	27% (69)
	4.1 ; +	CyO ; +
+ ; P4	29% (89)	18% (56)
+ ; +	28% (86)	25% (77)
	4.1 ; +	CyO ; +
+ ; P3	23% (57)	17% (42)
+ ; +	28% (69)	31% (76)
	6.5 ; +	CyO ; +
+ ; P1	9% (26)	18% (54)
+ ; +	45% (132)	27% (80)
	6.5 ; +	CyO ; +
+ ; P4	7% (12)	26% (42)
+ ; +	43% (69)	24% (39)
	6.5 ; +	CyO ; +
+ ; P3	13% (38)	22% (62)
+ ; +	41% (118)	24% (68)

Table 5.4. Combination of heterozygous DRhoGEF2 with heterozygous DMCC. For details of crosses see **section 2.16** of Materials and methods. *Drosophila* crosses were performed at 25°C and emerging adults scored from 10 days onwards until all had emerged. Percentages are the percentage number of flies of each genotype resulting from the cross. The expected percentage for each genotype is 25%. Numbers in parentheses are actual number of adult flies. P3 is a line with the same genetic background as P4 and is used here as a control (see **sections 4.21-4.24** of Chapter 4).

The combination of DRhoGEF2^{4.1} with a single DMCC allele does not lead to a decrease in viability compared to a wild-type allele (e.g. 26% versus 28% for P1) (**table 5.4**). There were no phenotypes observed for

this combination. Combining DRhoGEF2^{6.5} with DMCC does lead to a decrease in viability compared to wild-type (e.g. 9% versus 45% for P1).

	Px6 ; P1	CyO ; P1	Px6 ; +	CyO ; +
4.1 ; +	2% (4)	23% (38)	4% (6)	24% (40)
CyO ; +	23% (38)	0	24% (40)	0
	Px6 ; P4	CyO ; P4	Px6 ; +	CyO ; +
4.1 ; +	1% (2)	17% (34)	5% (10)	30% (59)
CyO ; +	17% (34)	0	30% (59)	0
	Px6 ; P3	CyO ; P3	Px6 ; +	CyO ; +
4.1 ; +	4% (8)	22.5% (60)	5% (12)	18% (40)
CyO ; +	22.5% (60)	0	18% (40)	0
	Px10 ; P1	CyO ; P1	Px10 ; +	CyO ; +
4.1 ; +	7% (12)	19% (33)	23% (40)	16.5% (29)
CyO ; +	19% (33)	0	16.5% (29)	0
	Px10 ; P4	CyO ; P4	Px10 ; +	CyO ; +
4.1 ; +	13% (34)	20.5% (53)	13% (34)	16% (41)
CyO ; +	20.5% (53)	0	16% (41)	0
	Px10 ; P3	CyO ; P3	Px10 ; +	CyO ; +
4.1 ; +	9% (16)	22% (39)	8% (14)	20% (36)
CyO ; +	22% (39)	0	20% (36)	0

Table 5.5. Combination of homozygous DRhoGEF2 with heterozygous DMCC. For details of crosses see **section 2.16** of Materials and Methods. *Drosophila* crosses were performed at 25°C and emerging adults scored from 10 days onwards until all had emerged. Percentages are the percentage number of flies of each genotype resulting from the cross. The expected percentage for each genotype is ~16.7% taking into account the homozygous lethality of CyO. Numbers in parentheses are actual number of adult flies. It was not possible to distinguish between the different DRhoGEF2 alleles, hence the total number for the two genotypes (e.g. 4.1/CyO ; P1/+ and Px6/CyO ; P1/+) was divided by two to give the number of adult flies and resulting percentage in each class. P3 is a line with the same genetic background as P4 and is used here as a control (see **sections 4.21–4.24** of Chapter 4).

However, the same is true when the control P3 allele is used (13% versus 41%), which indicates that it is not due to DMCC disruption. Again, there were no phenotypes observed. It is likely that there is another mutation/chromosomal aberration present on the P1, P3 and P4 chromosomes that is interacting with DRhoGEF2^{6.5} leading to a decrease in viability. It is also noticeable that the CyO balancer chromosome also gives a slight decrease in viability which leads to the >25% percentages for the other genotypes.

Given that there are 8 possible genotypes for each cross in **table 5.5** (homozygous DRhoGEF2 with heterozygous DMCC), each genotype would be expected to show a percentage viability of 12.5%. However, homozygous CyO is always lethal and therefore the percentage expected increases to ~16.7% for each genotype that survives. The combination of 4.1/Px6 leads to a significant decrease in viability with only 4-5% of flies surviving to adulthood. Given that the expected percentage of flies is 16.7% this is a percentage viability to adulthood of 26.9%, which fits with previous work that indicated the viability to adulthood of 4.1/Px6 is ~30% and 4.1/Px10 is 70% (Nikolaidou and Barrett 2004). Similarly, 4.1/Px10 has a decreased viability with an average of 14.5% of flies for that genotype surviving to adulthood which corresponds to a percentage viability to adulthood of 87%, which is rather higher than previously observed.

The addition of a DMCC allele to the combination of 4.1/Px6 does appear to decrease viability slightly when compared to the control (average 1.5% versus 4%). However, for the 4.1/Px10 combination the same was not true, with the viability of 4.1/Px10 in combination with P4 giving greater viability than with P3 (13% versus 9%), and with P1 giving a very similar figure to the P3 control (7% versus 9%). These differences could be merely due to the fact that the total number of adult flies for each experiment was not high (approximately 200). In order to see a statistical difference between two genotypes that produce very low numbers of

adults a much larger number of flies must be generated. The phenotypes observed in the flies that do make it to adulthood are no more or less severe with a DMCC allele than without (**fig. 5.5**).

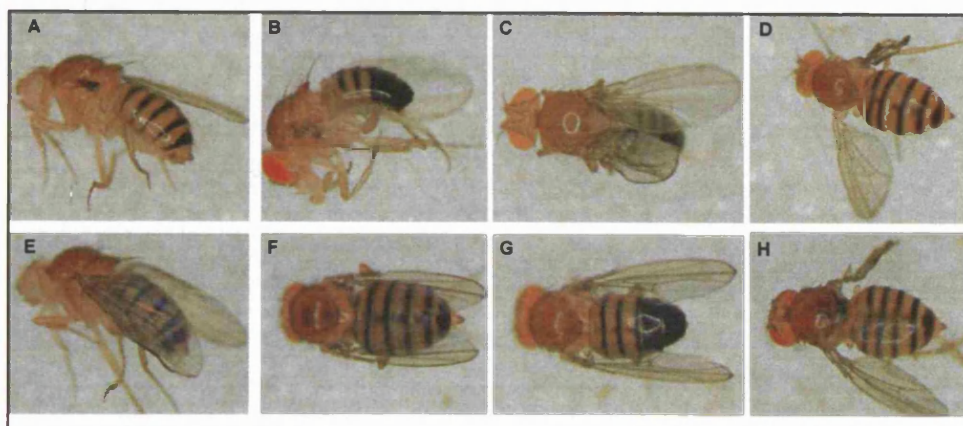


Figure 5.5. Phenotypes of flies homozygous mutant for DRhoGEF2 and heterozygous mutant for DMCC. For details of crosses see **section 2.16**, Materials and Methods. Adults were collected soon after emergence for imaging. Images here are representative of typical phenotypes observed. P3 is a line with the same genetic background as P4 and is used here as a control. **A:** 4.1/Px6 ; +/+, **B:** 4.1/Px6 ; P1/+, **C:** 4.1/Px6 ; P4/+, **D:** 4.1/Px6 ; P3/+, **E:** 4.1/Px10 ; +/+, **F:** 4.1/Px10 ; P1/+, **G:** 4.1/Px10 ; P4/+, **H:** 4.1/Px10 ; P3/P4.

The number of possible genotypes for the crosses in **table 5.6** is 16, therefore taking into account the homozygous lethality of CyO, the expected percentage viability for each genotype is ~8.3%. In this experiment the percentage viability of 4.1/Px6 ; +/+ ranged from 1 to 8%. 4.1/Px6 in combination with two DMCC alleles decreases this viability to zero. However, the control P3 allele in combination with either P1 or P4 and 4.1/Px6 is also not viable. Therefore, once again, the decrease in viability must be due to a mutation on the P1, P3 and P4 chromosomes other than DMCC.

The combination of 4.1/Px10 with two DMCC alleles leads to a similar decrease in viability to zero. However, one 4.1/Px10 ; P3/P4 fly did survive until adulthood, and this had a more severe wing phenotype than

is normally seen for 4.1/Px10 (fig. 5.5). Although this could indicate that MCC is genetically interacting with DRhoGEF2, crosses using much larger numbers of flies will need to be carried out to determine if there is a statistically significant decrease in viability as a result of DMCC alleles.

5.8. Discussion and conclusions

5.8.1. Molecular interaction between DRhoGEF2 and DMCC

DMCCa interacts with the DRhoGEF2 PDZ domain both in a yeast-2-hybrid experiment (K. Barrett, personal communication) and in an *in vitro* co-immunoprecipitation experiment (fig. 5.1). False positive interactions are common place in yeast 2-hybrid screens (Serebriiskii, Estojak et al. 2000). However, the fact that the recombinant DRhoGEF2 PDZ is able to pull DMCC from a cell lysate indicates that they are indeed able to interact. It also indicates that, if they do interact *in vivo*, the PDZ domain of DRhoGEF2 is sufficient for interaction.

Having determined that DMCC can interact with the DRhoGEF2 PDZ domain, the question then is why the two proteins did not give a positive interaction when full-length DRhoGEF2 was used (fig. 5.3). It is possible that within a “resting” cell the two proteins do not interact, but that, upon activation of a signalling pathway, they come together. This could be the pathway that likely activates DRhoGEF2 (see fig. 1.8 of Introduction), or a completely different pathway that feeds in to DRhoGEF2. DRhoGEF2 is a large protein, and it is possible that it folds into a conformation that shields the PDZ domain from interaction with targets until it receives a signal that enables it to adopt a binding conformation (fig. 5.6). Deletion of the C-terminus of any of the human orthologues of DRhoGEF2, p115RhoGEF, LARG and PDZ RhoGEF, increases their activity indicating that they may exist in an inactive form which could be the result of an intramolecular inhibition (see section 1.19 of Introduction). The human RhoGEFs homodimerise, and, although there is currently no evidence for this, it is possible that DRhoGEF2 does the same. This intermolecular interaction could prevent other interactors from binding.

	Px6 ; P1	CyO ; P1	Px6 ; +	CyO ; +
4.1 ; P4	0	9% (19)	3% (7)	14% (30)
CyO ; P4	9% (19)	0	14% (30)	0
4.1 ; +	2% (4)	10% (22)	4% (8)	13% (29)
CyO ; +	10% (22)	0	13% (29)	0
	Px6 ; P1	CyO ; P1	Px6 ; +	CyO ; +
4.1 ; P3	0	5% (9)	2% (4)	16% (27)
CyO ; P3	5% (9)	0	16% (27)	0
4.1 ; +	5% (8)	8% (14)	1% (2)	16% (28)
CyO ; +	8% (14)	0	16% (28)	0
	Px6 ; P3	CyO ; P3	Px6 ; +	CyO ; +
4.1 ; P4	0	8% (9)	3% (4)	11% (13)
CyO ; P4	8% (9)	0	11% (13)	0
4.1 ; +	3% (4)	11% (13)	8% (10)	13% (16)
CyO ; +	11% (13)	0	13% (16)	0
	Px10 ; P1	CyO ; P1	Px10 ; +	CyO ; +
4.1 ; P4	0	8% (14)	6% (10)	13% (23)
CyO ; P4	8% (14)	0	13% (23)	0
4.1 ; +	5% (8)	9% (16)	8% (14)	10% (18)
CyO ; +	9% (16)	0	10% (18)	0
	Px10 ; P1	CyO ; P1	Px10 ; +	CyO ; +
4.1 ; P3	0	5% (6)	7% (8)	13% (15)
CyO ; P3	5% (6)	0	13% (15)	0
4.1 ; +	11% (12)	7% (8)	11% (12)	9% (10)
CyO ; +	7% (8)	0	9% (10)	0
	Px10 ; P3	CyO ; P3	Px10 ; +	CyO ; +
4.1 ; P4	0.4% (1)	7% (18)	8% (20)	9% (22)
CyO ; P4	7% (18)	0	9% (22)	0
4.1 ; +	8% (20)	9% (22)	12% (30)	8% (20)
CyO ; +	9% (22)	0	8% (20)	0

Table 5.6. Combination of homozygous DRhoGEF2 with homozygous DMCC. (previous page). For details of crosses see **section 2.16, Materials and methods**. Percentages are the percentage number of flies of each genotype resulting from the cross. Numbers in parentheses are actual number of adult flies. The expected percentage for each genotype is ~8.3% taking into account the homozygous lethality of CyO. It was not possible to distinguish between the different DRhoGEF2 alleles nor to distinguish between P3 and P4, hence the total number for the two genotypes (e.g. 4.1/CyO ; P1/P4 and Px6/CyO ; P1/P4) or four genotypes (e.g. 4.1/CyO ; P3/+ and Px6/CyO ; P3/+ and 4.1/CyO ; P4/+ and Px6/CyO ; P4/+) was divided by two or four respectively to give the number of adult flies and resulting percentage in each class. P3 is a line with the same genetic background as P4 and is used here as a control.

Experiments were performed with phosphatase inhibitors in the lysis and immunoprecipitation buffers in order to test the possibility that DRhoGEF2 may exist in an inactive form to which DMCC cannot bind. Many signalling proteins are phosphorylated by kinases resulting in their activation, and dephosphorylated by phosphatases resulting in their inactivation. Human MCC becomes phosphorylated upon serum stimulation (Matsumine, Senda et al. 1996), and it is possible that the same is true for *Drosophila* MCC. The addition of phosphatase inhibitors, therefore, was an attempt to push the system towards a permanently active, signalling state. Similarly, the addition of GTPγS, which locks G proteins into an active signalling conformation, was another attempt to push the equilibrium towards a state of active signalling (with Concertina, the Gα protein that is likely to activate DRhoGEF2, particularly in mind). However, neither of these approaches led to a successful co-immunoprecipitation.

The co-immunoprecipitation experiments using full-length proteins were performed using *Drosophila* S2 cells which are macrophage-like (Ramet, Manfrulli et al. 2002). DRhoGEF2 has many potential signalling domains, and a signalling complex is likely to be built up around it. Some

of the factors contributing to this complex may be required in order for the PDZ domain to be active in binding targets, and these may not be expressed in S2 cells. Since these cells are derived from haemocytes which probably do not express DMCC (they were not seen to be stained in the *in situ*, chapter 3) and do not participate in epithelial folding events (which DRhoGEF2 is required for, see Introduction), they could contain factors inhibitory to interaction between DMCC and DRhoGEF2, or between DRhoGEF2 and other factors required for a “binding competent” PDZ domain. If this was the case, these factors would have to be inhibiting the interaction between DRhoGEF2 and DMCC via a site on DRhoGEF2 other than the PDZ domain, since the PDZ domain on its own can immunoprecipitate DMCC from an S2 cell lysate.

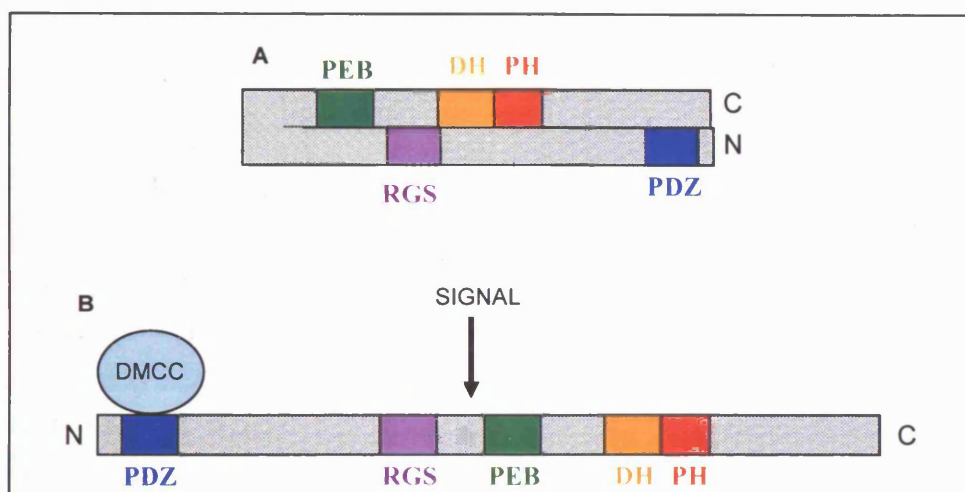


Figure 5.6. Model for DRhoGEF2 interaction with DMCC. **A:** DRhoGEF2 inactive conformation where the PDZ domain is inaccessible to target proteins. **B:** DRhoGEF2 changes to an active conformation as a result of an activating signal, and targets such as DMCC can bind. N: N-terminus, C: C-terminus.

Another factor that could influence the interaction is the choice of lysis buffer and wash buffer used for the immunoprecipitation. Lysis buffer contains detergent, and this can disrupt protein:protein interactions. For this reason a mild lysis buffer (20mM Tris-CL pH 7.5, m 1% NP40, 150mM NaCl) was used and this was also used as a wash buffer for the co-

immunoprecipitation experiments. It should be noted that the successful co-immunoprecipitation using the recombinant PDZ domain used these same buffers, therefore if these were influencing the protein:protein interaction it would have to be via a part of DRhoGEF2 other than the PDZ domain.

Perhaps the most likely explanation for the lack of interaction between DMCC and DRhoGEF2 when co-expressed in cells is that the expression levels are not sufficiently high to enable co-immunoprecipitation. Getting all three constructs (GAL4-Act5C, UAS-DRhoGEF2 and pFASTBAC1-DMCC) to express at levels high enough to visualise on a Western blot was difficult and unreliable. In many cases, a band was visualised on the whole cell lysate blot, but it was very weak indicating low expression levels. Even if all three were visualised on a blot, it was not necessarily the case that all three had co-transfected into any given single cell (although the fact that DRhoGEF2 was expressed indicates that it must be co-transfected with the Act-GAL4 construct). When attempting to co-transfect two constructs into the same S2 cell the success rate is >90% (personal communication, B. Baum), but for three constructs it is unknown. The rate of co-transfection of three constructs into a single cell is probably a significant limiting factor for these experiments. In order to overcome this, making use of the pFASTBAC1-Act5C vector to express DRhoGEF2, thereby negating the need for the GAL4-UAS system, could be employed. Alternatively, the DMCC construct could be expressed in one dish of cells and the Act5C and UAS-DRhoGEF2 constructs expressed in another, and the lysates of these could be mixed for the immunoprecipitation experiment.

In summary, although a successful co-immunoprecipitation of full-length DRhoGEF2 and DMCC was not achieved here, it does not necessarily mean that the two proteins do not interact within cells within a fruit fly due to the large number of factors potentially affecting the interaction in these experiments.

5.8.2. PDZ-target interaction specificity

The cell is a protein-rich environment, and at the region of the cell membrane there are likely to be a large number of PDZ proteins since many participate in signal transduction pathways at this cellular location. The *Drosophila* proteome contains 128 proteins (or predicted proteins) that contain PDZ domains (SMART search result). Many of these proteins contain more than one PDZ domain, and each PDZ domain may have a different set of targets to which it binds. Although not every PDZ-domain-containing protein will be expressed in every cell, specificity of binding is likely to be important, especially in the region of the cell membrane, to avoid potentially damaging cross-talk between signalling pathways.

DMCC carries a classic PDZ binding motif at its C-terminus: Threonine – x – Leucine – COOH. The number of potential target proteins with classic C-terminal PDZ-binding motifs (T/S/Y – X – V/L/I – COOH) in *Drosophila* is vast. Although the amino acids at the 0 and -2 positions of PDZ targets form specific interactions with residues of the PDZ binding pocket (Doyle, Lee et al. 1996), they are not the only amino acids important in the interaction. Structural studies have shown that the C-terminal five amino acids of the target all contribute to binding the PDZ domain (Skelton, Koehler et al. 2003). *In vitro* studies with peptide libraries indicate that perhaps as far back as the -8 position is important for binding (Songyang, Fanning et al. 1997). It is likely that PDZ-target interactions *in vivo* are determined by subtle differences in binding affinities of potential targets with PDZ domains based on the exact C-terminal sequence of the target. Thus, a given PDZ domain is likely to be able to bind to a subset of the large pool of potential targets based on this sequence specificity.

The fact that all three proteins identified in the yeast-2-hybrid with the DRhoGEF2 PDZ domain have a proline residue at the -4 position, as well as the common threonine at -2 and leucine at 0 positions, could be merely coincidence. However, all the yeast 2-hybrid positive interactors

that were later confirmed by co-immunoprecipitation have this sequence, whereas those that were not confirmed do not have PXTXL (they have DELEQ, IVTNL and VRTQN - K.Barrett, unpublished). Therefore, it is likely that this sequence is important for binding to DRhoGEF2. It should be noted that, recently, a fourth potential interactor for DRhoGEF2 has been identified by yeast 2-hybrid. This interactor carries an IXTXL motif at its C-terminus (M.Leptin, unpublished).

The results of the mutation analysis (fig. 5.2) illustrate that the proline is likely to be necessary for binding to DRhoGEF2 since mutation of this residue to an alanine inhibits binding. Similarly, mutation of the leucine residue at the 0 position to an alanine prevents the PDZ domain from binding to DMCC. It should be noted that the total amount of immunoprecipitated protein in the DMCC^{L→A} lane did not appear to be as high as for the other mutants, since the background bands in the Western blot are fainter than in the other lanes (fig. 5.2 blot A). The expression levels for this mutant appeared to be the same as the other mutants in the whole cell lysate analysis (fig. 5.2 blot C), and therefore it appears there was an inefficiency in immunoprecipitation in this particular case.

Mutation of the threonine residue to an alanine compromises binding of DMCC to DRhoGEF2, but does not prevent it altogether. Although it is tempting to speculate that this implies the threonine is not as important for binding DRhoGEF2 as the proline or leucine, it would be important to quantify the interaction, for example using a fluorescein-tagged PXTXL peptide in a colorimetric assay (Lim, Hall et al. 2002). Although the co-immunoprecipitation experiment is not highly quantitative, it does illustrate that the proline, threonine, and leucine residues are likely to be important for the interaction between DMCC and DRhoGEF2. It should be kept in mind that these experiments were not performed with full-length DRhoGEF2, and, as described in section 5.9.1, there are other factors that must be taken into account when considering whether these interactions actually take place *in vivo*. It should also be noted that these

results contradict previous results using the same MCC mutants in a yeast 2-hybrid approach. In this case, both the DMCC^{T→A} and DMCC^{L→A} mutants showed a complete loss of interaction, whereas the DMCC^{P→A} mutation had no effect on the interaction (K.Barrett, unpublished).

The DMCCd isoform also co-immunoprecipitated with DRhoGEF2 PDZ domain. This isoform is truncated and lacks the C-terminal PXTXL motif. This result implies that this motif is not required for binding to DRhoGEF2 PDZ domain, and presumably therefore, that the binding site is elsewhere within DMCC. However, this protein was vastly over-expressed compared to the wild-type and mutants (fig. 5.2 blot C) for reasons that are unclear. It is possible that an MCC isoform lacking the C-terminus is more stable than full-length MCC. The extremely high concentration of this particular isoform could lead to it artificially binding to the PDZ domain. It is unlikely that this isoform does bind the PDZ domain *per se* since the full-length proteins with a single amino acid change at the C-terminus do not bind whereas the wild-type does, indicating that the binding site is highly likely to be at the C-terminus.

It is possible that amino acids other than proline, threonine or leucine at the -4, -2 and 0 positions respectively could also allow binding of DMCC to DRhoGEF2, and that these amino acids are not strictly necessary for binding since others could perhaps substitute. In order to test this, the complete complement of amino acids would have to be tested at the 0, -2 and -4 positions in an interaction assay. It is likely that most amino acids would not successfully substitute due to the nature of the specific interactions formed between a PDZ domain and its target.

In order to test whether the proline, threonine and leucine residues are sufficient for binding to the DRhoGEF2 PDZ domain, the PXTXL motif could be introduced to the C-terminus of a protein that does not carry such a motif (and does not bind DRhoGEF2) and this protein tested for binding to the PDZ domain of DRhoGEF2. As before, each amino acid

could be mutated to an alanine in turn to determine which combination is sufficient for binding. Ideally, these experiments would be performed with full-length DRhoGEF2 co-transfected with the mutants into cells in order to make the experiment as physiologically relevant as possible.

In order to determine whether the proline at the -4 position (in combination with the TXL motif) confers specificity of binding to DRhoGEF2 there are two questions that could be asked: are proteins with C-terminal motifs other than PXTXL able to bind DRhoGEF2, and are proteins with PXTXL motifs at their C-terminus able to bind to other PDZ domains?

To answer the first of these questions, DMCC (or another of the PXTXL proteins) could be co-transfected into cells with a protein containing a PDZ domain that does not preferentially select for PXTXL, but does select for XXXTL (such as Shank – see below), and co-immunoprecipitation tests performed. Ideally, a number of different XXXTL-selecting PDZ domains would be tested. If DMCC did bind to these other PDZ domains it would indicate that the proline does not allow exclusive binding to DRhoGEF2 for PXTXL proteins. To test the second question, a protein containing a ZXTXL C-terminal motif, where Z is an amino acid known to be selected for by a different PDZ domain, could be tested in a similar co-immunoprecipitation experiment using the DRhoGEF2 PDZ domain. If PDZ target proteins carrying non-PXTXL motifs were able to bind to DRhoGEF2 this would indicate that the proline was not necessary for binding DRhoGEF2. The results of both experiments would carry more weight if carried out in conjunction with a quantitative *in vitro* approach such as a colorimetric assay.

Specificity for a proline at the -4 position *in vivo* has not been documented for a PDZ domain to date. The third PDZ domain of human Ptp-bas3 strongly selects for a proline at this position in an *in vitro* experiment using a peptide library (Songyang, Fanning et al. 1997).

However, *in vivo* the third PDZ domain of Ptp-bas3 interacts with the Fas antigen which has the C-terminal sequence IQSLV-COOH.

5.8.3. Modelling PDZ domain interactions

In order to reliably model the interaction of a PDZ domain with a potential target peptide, the crystal structure of the protein is required. For this reason, it is not possible to model the PDZ domain of DRhoGEF2 with any reliability. The PDZ domain of the human protein Shank shows selectivity for targets with TXL at their C-terminus. A known target of the Shank1 PDZ domain is Guanylate kinase associated protein (GKAP1), which has an EAQTRL motif at its C-terminus.

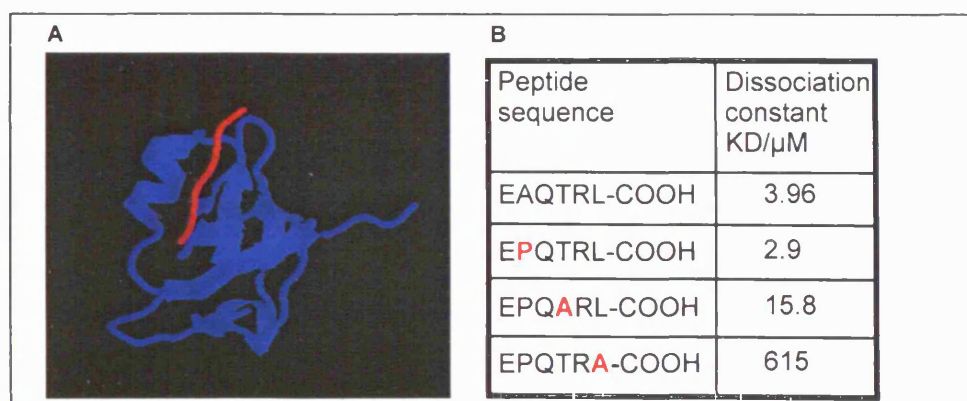


Figure 5.7. Model of the Shank PDZ domain interacting with a peptide target. **A:** Ribbon model of the Shank PDZ domain (blue) interacting with the peptide EAQTRL-COOH (red). **B:** predicted dissociation constants for the Shank:peptide interaction. For details of the methods see (Reina, Lacroix et al. 2002). Model generated by G. Fernandez. EMBL.

The crystal structure of this interaction has been solved (Im, Lee et al. 2003). This crystal structure was used as the basis for a computational model of the Shank PDZ domain binding to a short peptide (fig. 5.7, G.Fernandez). This computational model makes use of novel computer-aided design techniques (Reina, Lacroix et al. 2002).

Using the model, predictions were run to test the same C-terminal sequences used in the co-immunoprecipitation experiments. The results predict that mutation of the leucine has a strong effect on binding to the PDZ domain, mutation of the threonine has some effect, and mutation of the proline has very little effect. The results of this computational prediction may indicate that the Shank PDZ domain is not selective for the amino acid at the -4 position since the dissociation constant is predicted to be similar for an alanine residue at the -4 position (the “wild-type”) and a proline residue. It would be very interesting to perform such a prediction using the PDZ domain of DRhoGEF2, but unfortunately this is not presently possible due to the lack of crystal structure.

In order to predict the amino acids of the DRhoGEF2 PDZ domain that may be interacting with the 0, -2 and -4 positions of its targets, the best that can currently be done is to line up the primary sequence of the DRhoGEF2 PDZ domain with PDZ domains whose crystal structures have been solved. The third PDZ domain of psd-95 was the first crystal structure of a PDZ domain to be solved (Doyle, Lee et al. 1996). The C-terminal residue of the psd-95 PDZ3 target interacts with the characteristic Glycine-Leucine-Glycine-Phenylalanine (GLGF) motif seen in many PDZ domains (fig. 5.8). The -2 position residue of the target interacts with the residue immediately following the GLGF motif, and the histidine at the base of the α B alpha helix. The -4 position does not appear to be significantly involved in binding, and lies outside the PDZ pocket.

Lining up the primary sequence of the DRhoGEF2 PDZ domain with the psd-95 PDZ domain illustrates that the GLGF motif of psd-95 is a GYGM motif in DRhoGEF2, and that the histidine at the base of the α B helix of psd-95 is conserved in DRhoGEF2 (fig. 5.9). The GYGM motif is likely to interact with the C-terminal leucine of the DRhoGEF2 predicted PXTXL targets, and the conserved histidine of DRhoGEF2 with the threonine residue at the -2 position.

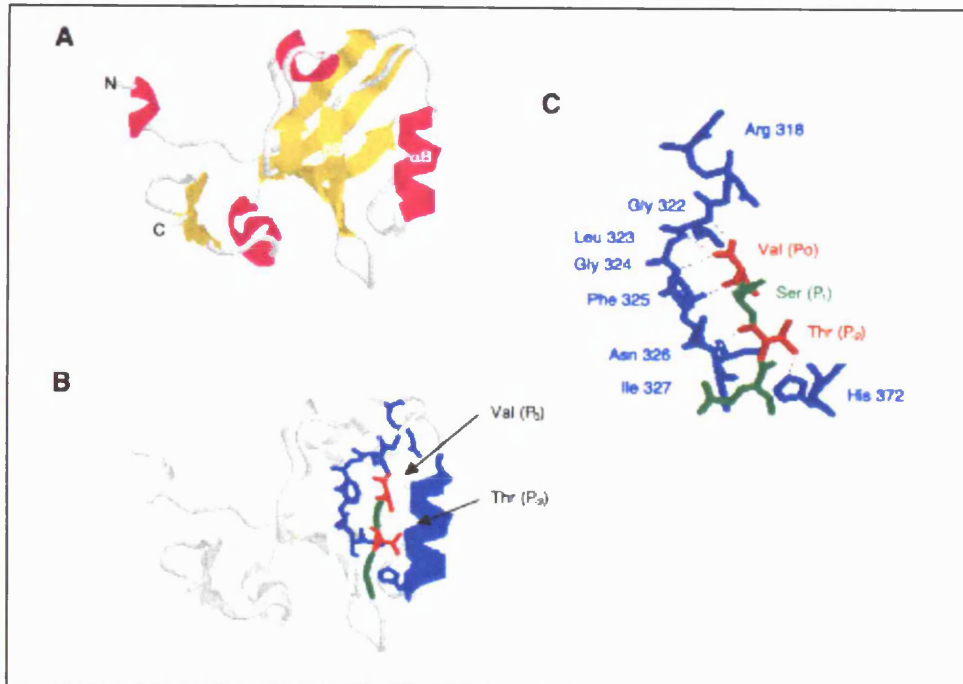


Figure 5.8. PDZ-target interactions. Crystal structure of the binding pocket of psd-95 PDZ3 interacting with a peptide target. **A:** PDZ domain ribbon structure, **B and C:** PDZ binding pocket (blue), target peptide (red/green). Taken from (Nourry, Grant et al. 2003)

Comparing DRhoGEF2 PDZ domain with the Shank PDZ domain, which binds a target with an AQTRL motif at its C-terminus, and the FAP-1 PDZ domain, which selects for a proline at the -4 position in an *in vitro* assay (Songyang, Fanning et al. 1997), one can see that there is no more similarity between these proteins and DRhoGEF2 than there is between DRhoGEF2 and psd-95 PDZ3 within the regions that interact with the C-terminus of the target for psd-95. This indicates that for each PDZ domain there are likely to be additional or different amino acids of the PDZ domain forming specific interactions with the target than those indicated in yellow in **figure 5.9**. The crystal structure is therefore an invaluable tool in predicting PDZ binding specificity.

The amino acid proline has a small, non-polar side chain and is mildly hydrophobic (as glycine and alanine). It is not an amino acid commonly associated with protein-protein interaction sites, and is more likely to be found at the loops or bends of proteins. It could, therefore, be the case that the proline residue at the -4 position of DRhoGEF2 predicted targets, is not involved *per se* in binding to DRhoGEF2, but instead is required to provide a loop in the C-terminus of the target that enables the formation of specific interactions between amino acids of the target and the PDZ binding pocket or residues around it.

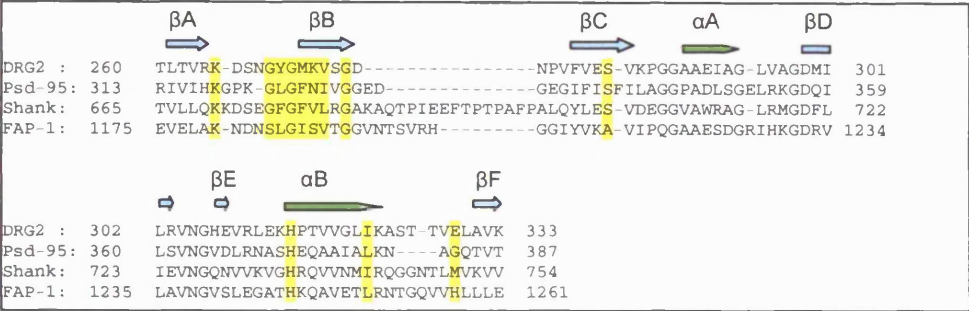


Figure 5.9. PDZ domain primary sequence line-ups. DRG2: DRhoGEF2 PDZ domain, psd-95: psd-95 PDZ3 domain, Shank; Shank PDZ domain, FAP-1: FAP-1 PDZ domain. Yellow highlights residues that are important in forming interactions with the target peptide for psd-95 PDZ3. Blue arrows indicates beta sheets, green arrows indicate alpha helices. Numbers refer to published amino acid sequences of the full-length proteins.

5.8.4. Additional targets

If the DRhoGEF2 PDZ domain does indeed select for a PXTXL motif *in vivo*, there are three further potential targets in the *Drosophila* proteome (table 5.2) in addition to the three identified with the yeast 2-hybrid experiment. The approach of determining the consensus binding sequence for a particular PDZ domain and subjecting this sequence to a genomic search to identify potential targets has been carried out previously. The consensus binding sequences (the last five amino acids) for the second and third PDZ domains of psd-95 were determined, and these used to BLAST search the C-terminal human proteome revealing

fifty potential targets (Lim, Hall et al. 2002). Three of these (with available antibodies) were tested by co-immunoprecipitation and two were found to bind. This novel method of identifying targets for PDZ domains is one that may become popular in this age of availability of genomic sequence, although it is likely to reveal false positives due to the consensus sequence being of inadequate stringency, and identify many interactions that could never physically happen *in vivo* due to non-overlapping expression patterns.

The following section describes analysis of the available data for the three DRhoGEF2 PDZ targets identified in the yeast 2-hybrid and the three potential targets identified bioinformatically based on their C-terminal sequence. There is currently no evidence that the latter three genes have any link to DRhoGEF2, and therefore the proposed experiments would be highly speculative. For the three targets identified in the yeast 2-hybrid and confirmed by immunoprecipitation there is biochemical evidence of an interaction, but this has not been functionally confirmed for any of the three to date.

The three potential DRhoGEF2 targets identified by proteome-wide BLAST analysis could be tested in a co-immunoprecipitation experiment with DRhoGEF2. It is interesting to note that all three of these proteins contain predicted signal sequences at their N-termini implying that they are processed along the secretory pathway. The yeast 2-hybrid screen using the DRhoGEF2 PDZ domain was performed using an S2 cell library with the GAL4 activation domain and nuclear localisation signal placed at the N-terminus of the protein (Du, Vidal et al. 1996). It is unlikely, therefore, that an N-terminal signal sequence would have been successful in sending the protein up the secretory pathway and not to the nucleus where the assay is performed, but this could perhaps provide an explanation for why these targets were not picked up in the screen if they are indeed real targets.

Of the three, one has been characterised in *Drosophila*, this being capricious (caps). Caps protein is expressed in subsets of neurons and muscles during late stages of embryogenesis, and facilitates the correct partnering of a motor neuron with its target muscle (Shishido, Takeichi et al. 1998). Caps is also required for proper dorsal/ventral boundary formation during *Drosophila* imaginal disc development (Milan, Weihe et al. 2001). In this case, caps protein is observed in filopodia-like structures that extend from ventral cells over the surface of dorsal cells at

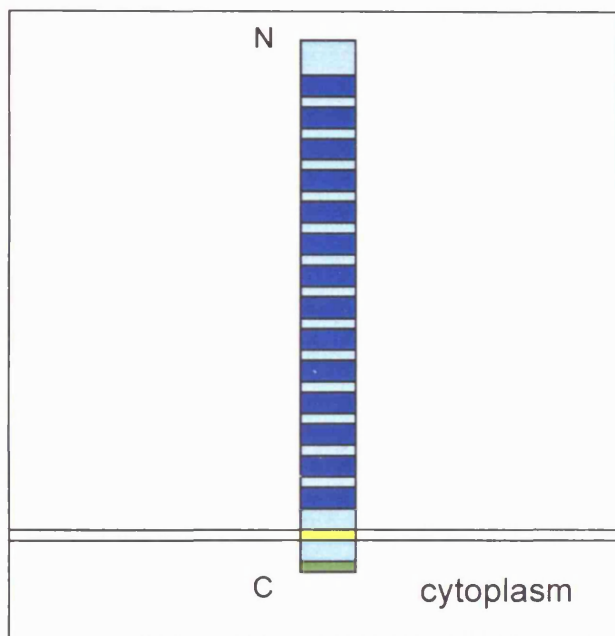


Figure 5.10. Caps protein structure. Dark blue indicates leucine-rich repeats, yellow indicates transmembrane domain, green indicates PXTXL at C-terminus. N: N-terminus, C: C-terminus. Not to scale.

the dorsal/ventral boundary (Milan, Weihe et al. 2001). Caps is a transmembrane protein with fourteen leucine rich repeats (LRR) in its extracellular domain (**fig. 5.10**). LRR domains may mediate protein:protein interactions (Kobe and Deisenhofer 1994), and although it is tempting to speculate that caps functions by homophillic cell:cell interactions, this does not appear to be the case since S2 cells do not aggregate if caps

is expressed on their surface (Shishido, Takeichi et al. 1998; Milan, Weihe et al. 2001). Thus, it is likely that there is an, as yet undiscovered, cell surface binding partner of caps.

In order for any of the six PXTXL targets to interact with DRhoGEF2, they would have to be expressed in the same tissue, be localised to the same

sub-cellular region, and be expressed at the same stage of development. DRhoGEF2 mRNA is maternally loaded and expressed at a low level almost ubiquitously throughout embryogenesis (Barrett, Leptin et al. 1997). DRhoGEF2 protein is expressed predominantly in epithelia throughout embryogenesis (Padash Barmchi, Rogers et al. 2005). DRhoGEF2 protein is specifically enriched in the furrow canals during cellularisation (stage 5) (Grosshans, Wenzl et al. 2005), in the ventral furrow during gastrulation (stage 6/7), in the lateral epidermis during dorsal closure (stage 13), in ventral epidermal cells of thoracic and abdominal segments in a repeated pattern (stage 14), and in the longitudinal and commissural axons of the central nervous system (stage 17) (Padash Barmchi, Rogers et al. 2005).

Although there is no data on the protein localisation of the PXTXL genes (with the exception of capricious), there is information on the mRNA localisation for some. DMCC mRNA is expressed in the central nervous system (**fig 5.12** of Chapter 3), group 3 mRNA is maternally loaded and expressed in the germ band, central nervous system and gut (G. Escott, unpublished), Capricious protein is expressed in a subset of motoneurons and muscles of late stage embryos (Shishido, Takeichi et al. 1998), and CG3777 mRNA is expressed in the posterior spiracles, the tracheal system, the salivary glands and the dorsal and ventral epidermis from around stage 11 of embryogenesis (data from the BDGP *in situ* database website - <http://www.fruitfly.org/cgi-bin/ex/insitu.pl>). If the mRNA localisations reflect protein localisation, DMCC, Group 3, Caps and CG3777 proteins could all, potentially, interact with DRhoGEF2 since their expression patterns overlap.

Although there is no *in situ* data available for the other PXTXL proteins, there is microarray data which catalogues their expression during embryogenesis (**fig. 5.11**). The microarray data is of limited use since it gives no information about the tissues in which the gene is expressed. However, it does allow comparison of expression levels at one stage of

embryogenesis with another. **Figure 5.11** illustrates that all the PXTXL genes are expressed during embryogenesis.

The microarray data suggests that DRhoGEF2 mRNA has a particular expression level pattern during embryogenesis. DRhoGEF2 mRNA levels are particularly high at around stages 2-4 before they drop off and then rise to a second high at around stage 11-12 where they persist until the end of embryogenesis (**fig. 5.11**). None of the PXTXL genes have this exact pattern of expression levels, but one has highest expression levels early in embryogenesis from around stages 1-5 (Group 3) and four have highest expression levels around stage 13 (Mec-2, MCC, caps and CG3777). The microarray data, therefore, illustrates that DRhoGEF2 is expressed at high levels at the same time that many of the PXTXL genes are also experiencing high expression levels.

At the subcellular level, DRhoGEF2 is localised apically in cells of the invaginating ventral furrow and later (stage 11) is apically localised in all epidermal cells (Padash Barmchi, Rogers et al. 2005). In S2 cell culture DRhoGEF2 co-localises with the plus ends of microtubules (Rogers, Wiedemann et al. 2004). Given that DRhoGEF2 interacts with Rho, which, by virtue of its geranylgeranylation, is localised to the plasma membrane upon activation, and that DRhoGEF2 probably interacts with Cta, which is likely to be plasma membrane localised upon activation of its G-protein coupled receptor presumptive binding partner, it is likely that DRhoGEF2 is also recruited to the plasma membrane upon activation of signalling through the putative pathway from Cta to Rho via DRhoGEF2. DRhoGEF2 would be an unusual PDZ-containing protein if it were not localised in the region of the plasma membrane when it interacted with its PDZ targets.

There is no sub-cellular localisation information available for any of the PXTXL genes except Caps which is localised to the synaptic sites of both motor neurons and muscles (Shishido, Takeichi et al. 1998).

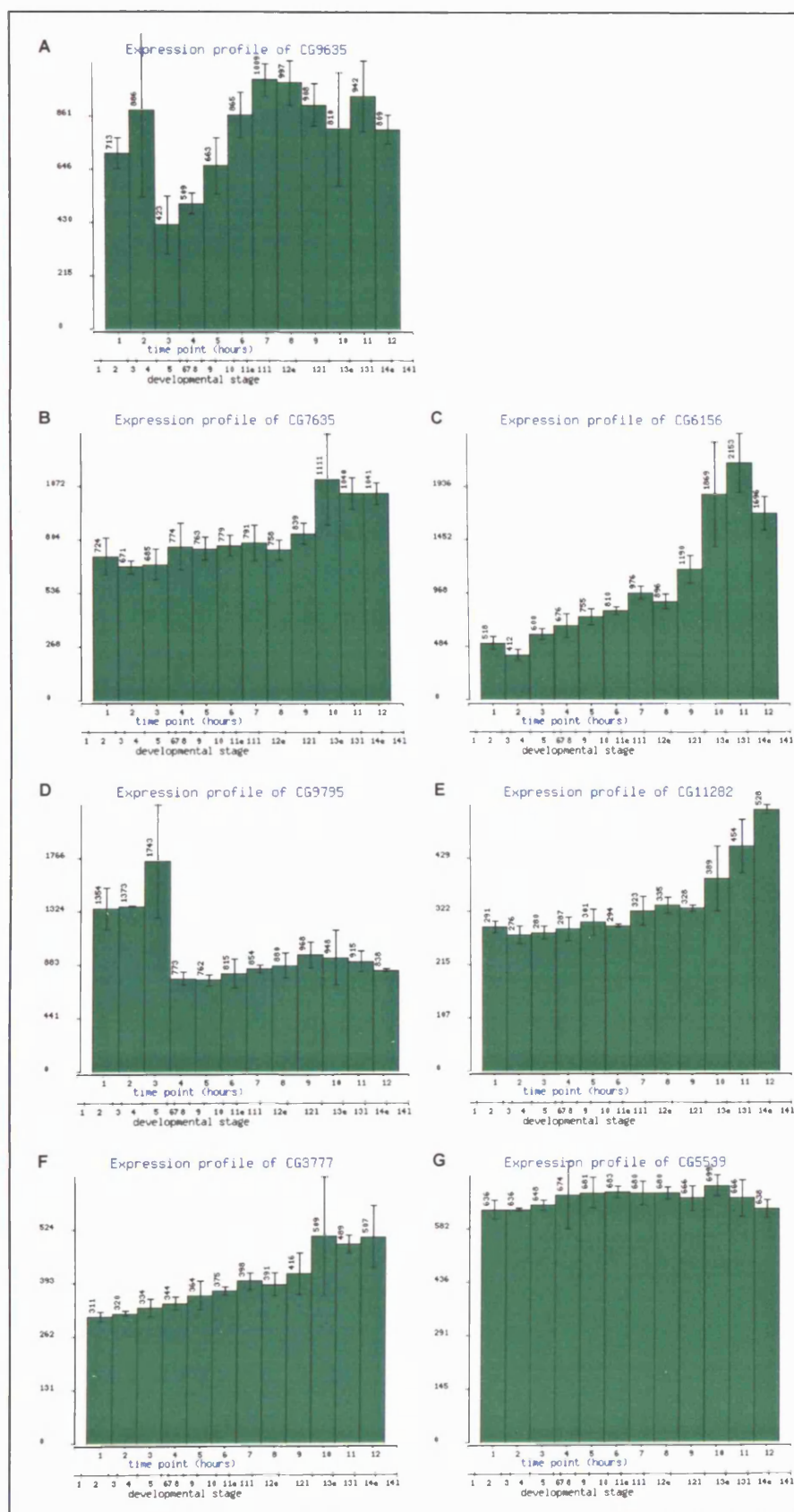


Figure 5.11. Embryonic gene expression profiles of DRhoGEF2 and the six PXTXL genes (previous page). Microarray expression data taken from the BDGP expression database website - <http://www.fruitfly.org/cgi-bin/ex/insitu.pl>. For details of methods see **fig. 3.11**, chapter 3) **A:** DRhoGEF2, **B:** Mec-2, **C:** DMCC, **D:** Group 3, **E:** Caps, **F:** CG3777, **G:** CG5539. Green indicates the mRNA is present, blue indicates marginal, red indicates absent. The y axis cannot be compared between different graphs as it represents an arbitrary unit.

Of the six potential *in vivo* binding partners for the DRhoGEF2 PDZ domain, caps has a single predicted transmembrane domain, and CG5539 predicted protein has two predicted transmembrane domains. A signal sequence plus a single transmembrane domain leads to the topology illustrated in **figure 5.10**, where the N-terminus is extracellular (or luminal if the protein is not at the cell surface), and the C-terminus is cytoplasmic. The existence of a second transmembrane domain indicates that the C-terminus is likely to be extracellular/luminal. In this respect, it is possible that Caps C-terminus could interact with DRhoGEF2 PDZ domain, but unlikely that CG5539 predicted protein C-terminus could. CG3777 protein has a signal sequence but no predicted transmembrane domain, and may, therefore, be secreted or directed to the lysosome. It is unlikely, therefore, that it would interact with DRhoGEF2 within a cell since DRhoGEF2 is likely to be cytoplasmically localised. Therefore, of the three potential PXTXL targets identified by proteome search, only Caps protein is likely to be able to interact with DRhoGEF2 via its C-terminus. None of the three targets identified in the yeast 2-hybrid have a signal sequence or predicted transmembrane domains, and therefore their localisation would be presumed to be cytoplasmic, and they would therefore be able to interact with DRhoGEF2.

Of particular interest would be how the targets compete for interaction with the DRhoGEF2 PDZ domain if they are expressed in the same cell. Presumably only one target molecule could interact with one PDZ domain

at any one time, and it may be the case that the target with the highest affinity target sequence would “win” for interaction (in this case the X residues of PXTXL may determine subtle differences in the binding affinity between targets). Alternatively, there may be subcellular localisation mechanisms for the targets that allow them to interact with different pools of DRhoGEF2 in different regions of the cell and therefore competition between targets would not be relevant.

5.8.5. Genetic interactions

In order to better understand the interaction between DRhoGEF2 and DMCC, genetic interaction experiments were performed to look for suppression or enhancement of a DRhoGEF2 phenotype by MCC mutant alleles.

The results of the genetic interaction experiments indicate that there is not a strong genetic interaction between DRhoGEF2 and DMCC alleles. There does appear to be a weak genetic interaction observed in the experiments, but this is likely to be a result of mutations on the chromosome other than DMCC. The implication of these results is that DMCC does not have a significant effect on DRhoGEF2 function during development since mutations in DMCC do not enhance or suppress the DRhoGEF2 phenotype. It is, therefore, likely that DMCC is not participating in the signalling pathway governing cell shape changes in which DRhoGEF2 functions. Since there were no additional phenotypes observed which could not be attributed to DRhoGEF2 function, it is not possible to say anything further about DMCC function.

5.9. Chapter 5 References

Barrett, K., M. Leptin, et al. (1997). "The Rho GTPase and a putative RhoGEF mediate a signaling pathway for the cell shape changes in *Drosophila* gastrulation." Cell **91**(7): 905-15.

- Doyle, D. A., A. Lee, et al. (1996). "Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ." Cell **85**(7): 1067-76.
- Du, W., M. Vidal, et al. (1996). "RBF, a novel RB-related gene that regulates E2F activity and interacts with cyclin E in Drosophila." Genes Dev **10**(10): 1206-18.
- Grosshans, J., C. Wenzl, et al. (2005). "RhoGEF2 and the formin Dia control the formation of the furrow canal by directed actin assembly during Drosophila cellularisation." Development **132**(5): 1009-1020.
- Im, Y. J., J. H. Lee, et al. (2003). "Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization." J Biol Chem **278**(48): 48099-104.
- Kobe, B. and J. Deisenhofer (1994). "The leucine-rich repeat: a versatile binding motif." Trends Biochem Sci **19**(10): 415-21.
- Lim, I. A., D. D. Hall, et al. (2002). "Selectivity and promiscuity of the first and second PDZ domains of PSD-95 and synapse-associated protein 102." J Biol Chem **277**(24): 21697-711.
- Matsumine, A., T. Senda, et al. (1996). "MCC, a cytoplasmic protein that blocks cell cycle progression from the G0/G1 to S phase." J Biol Chem **271**(17): 10341-6.
- Milan, M., U. Weihe, et al. (2001). "The LRR proteins capricious and Tartan mediate cell interactions during DV boundary formation in the Drosophila wing." Cell **106**(6): 785-94.
- Nikolaidou, K. K. and K. Barrett (2004). "A Rho GTPase Signaling Pathway Is Used Reiteratively in Epithelial Folding and Potentially Selects the Outcome of Rho Activation." Curr Biol **14**(20): 1822-6.
- Nourry, C., S. G. Grant, et al. (2003). "PDZ domain proteins: plug and play!" Sci STKE **2003**(179): RE7.
- Padash Barmchi, M., S. Rogers, et al. (2005). "DRhoGEF2 regulates actin organization and contractility in the Drosophila blastoderm embryo." J Cell Biol.
- PATSCAN "<http://www.flybase.net/patscan/>."

- Ramet, M., P. Manfrulli, et al. (2002). "Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*." Nature **416**(6881): 644-8.
- Reina, J., E. Lacroix, et al. (2002). "Computer-aided design of a PDZ domain to recognize new target sequences." Nat Struct Biol **9**(8): 621-7.
- Rogers, S. L., U. Wiedemann, et al. (2004). "*Drosophila* RhoGEF2 associates with microtubule plus ends in an EB1-dependent manner." Curr Biol **14**(20): 1827-33.
- Serebriiskii, I., J. Estojak, et al. (2000). "Approaches to detecting false positives in yeast two-hybrid systems." Biotechniques **28**(2): 328-30, 332-6.
- Shishido, E., M. Takeichi, et al. (1998). "*Drosophila* synapse formation: regulation by transmembrane protein with Leu-rich repeats, CAPRICIOUS." Science **280**(5372): 2118-21.
- Skelton, N. J., M. F. Koehler, et al. (2003). "Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain." J Biol Chem **278**(9): 7645-54.
- Songyang, Z., A. S. Fanning, et al. (1997). "Recognition of unique carboxyl-terminal motifs by distinct PDZ domains." Science **275**(5296): 73-7.

6. Summary, conclusions and future work

The hypothesis behind the work in this thesis was:

The *Drosophila melanogaster* orthologue of human MCC interacts with DRhoGEF2 during development, and this interaction affects DRhoGEF2 signalling and *Drosophila* development.

This thesis set out with two main goals: to characterise the function of *Drosophila* MCC, and investigate the significance of its interaction with DRhoGEF2.

6.1. Summary of results

In chapter 3 the *Drosophila* orthologue of the human MCC gene was investigated and confirmation presented that only one orthologue exists in *Drosophila*. Although the *Drosophila* predicted protein bears limited primary sequence similarity to the human protein (up to 51% in regions of high similarity), the likely functional motifs are conserved i.e. five coiled coils, a potentially significant arginine residue, and a C-terminal PDZ binding motif. BLAST analysis revealed the presence of MCC orthologues in various species, many carrying a C-terminal PDZ-binding motif. mRNA expression analysis revealed that there are likely to be at least four different mRNAs produced from the MCC locus, two of which include an exon from a neighbouring predicted gene. mRNA expression analysis during development revealed that MCC is expressed predominantly in the central nervous system.

In chapter 4 a preliminary analysis of loss-of-function and gain-of-function MCC scenarios, concluded that neither has an obviously visible effect on development. A further analysis of function of MCC was pursued by making an MCC mutant. Two P-element insertion lines were generated, the first an insertion and deletion that disrupts other genes in addition to MCC, the second a clean insertion into MCC. The MCC mutant shows

no obviously visible phenotype, and is viable when combined with a deficiency removing the gene.

Results from chapter 5 demonstrate that DMCC can interact with the isolated PDZ domain of DRhoGEF2, but it was not shown to interact with the full-length protein. Genetic interaction analysis revealed that DMCC mutation does not have a significant impact on the DRhoGEF2 mutant phenotype, indicating that DMCC is not likely to participate in a genetic pathway with DRhoGEF2.

6.2. MCC function

Overall, the results disprove the hypothesis that DMCC affects DRhoGEF2 signalling and *Drosophila* development since DMCC does not genetically interact with DRhoGEF2, and does not itself have a phenotype when mutated or over-expressed. The lack of phenotype upon MCC mutation was unexpected, and raises the question as to the function of DMCC in *Drosophila*. There are two possibilities, the first of which breaks down into two categories:

- MCC has a function in *Drosophila*:
 - that is redundant with the function of another gene
 - that is more subtle than the analysis presented here can detect
- MCC has no function in *Drosophila*

The possibility that MCC has functional redundancy with another gene is quite likely. Although there is no MCC primary sequence homologue in flies, there may be a functional homologue i.e. there could be another protein expressed that is able to compensate for loss of MCC function in an MCC mutant. Therefore, a phenotype would only be observed when both genes were mutated. Since MCC protein is likely to be

predominantly localised in the central nervous system and possibly the gut, this protein would presumably function in these tissues.

If MCC acts as a tumour suppressor in flies, as it is proposed to do in humans, it is possible that an effect of MCC mutation could be seen in a background of perturbation of other genes promoting overgrowth. Human cancer patients with mutated MCC have mutations in many other genes that contribute to the stepwise generation of colon cancer (Fearon and Vogelstein 1990). It is possible that MCC mutation in the fruit fly would contribute to tumour formation if additional mutations accumulated, but that the lifetime of a fly is too short for this to occur. Since MCC is predominantly expressed at high levels in the central nervous system of the fly, and not in various other epithelia as it is in the mouse, any phenotype associated with loss of tumour suppressor function may not be seen in the fly (since the majority of cells of the nervous system are not continuously growing and dividing like epithelial cells do).

It is possible that MCC mutation gives rise to a subtle phenotype that is not observed here but is present. Since MCC protein is likely to be functioning in the central nervous system this phenotype would presumably be manifested there. This phenotype would probably mildly affect the development and/or function of the cells in which MCC is functioning, whether they are neurons or other cells of the central nervous system. However, it would presumably not grossly affect the function of the CNS since this would have been observed as a phenotype. Although MCC mutant flies appear to feed and behave normally, they may exhibit a subtle phenotype if subjected to behavioural tests for example.

The final possibility is that MCC has no function in the fly. It is hard to imagine that the expression of a protein for which there is no function would persist during evolution. However, there are proteins that are expressed for which no function is known, the mammalian prion protein being an example. Prion knockouts have no phenotype and there has

been no cellular function assigned to the prion protein, yet prion is expressed in various mammalian tissues (Prusiner 1998).

Since the MCC mutant in *Drosophila* has no obvious phenotype, but MCC appears to provide a tumour suppressor role in mammals, it would be interesting to characterise the function of mammalian MCC further.

Experiments have confirmed the role proposed by Matsumine *et al* of MCC as providing a block on the cell cycle at the G1 to S phase transition (S.Loh, unpublished). Preliminary evidence suggests that MCC may play a role in the regulation of the actin cytoskeleton, since MCC over-expression leads to a mild increase in stress fibre formation in NIH3T3 cells (S.Loh, unpublished). PDZRhoGEF over-expression leads to a redistribution of actin to the cell cortex and cell rounding (Togashi, Nagata et al. 2000; Banerjee and Wedegaertner 2004). Over expression of PDZRhoGEF leads to Rho activation using an SRE reporter assay, and this is further stimulated by concurrent over-expression of MCC (L. Gardano, unpublished). A link between MCC, Rho and the actin cytoskeleton could indicate that MCC modulates RhoGEF's activation of Rho.

It would be interesting to test whether mutation of the arginine residue that appears to be important in cell cycle control has any effect on MCC's regulation of Rho or the actin cytoskeleton. Since mammalian RhoA is involved in the regulation of entry into G1 and progression into S-phase (Olson, Ashworth et al. 1995), it is possible that MCC exerts its influence on cell cycle progression via Rho.

6.3. DRhoGEF2 PDZ domain function

The lack of genetic interaction between DRhoGEF2 and DMCC also raises the question as to the function of the PDZ domain of DRhoGEF2. The PDZ domain of DRhoGEF2 is conserved in the human orthologues, PDZRhoGEF and LARG, and in these orthologues it is used as a protein binding module, for example for binding to plexin B (Swiercz, Kuner et al. 2002). Unless the PDZ domain of DRhoGEF2 is very unusual, it is

probable that its function is to mediate interaction with other proteins. The interaction of a protein with the PDZ domain of DRhoGEF2 could have two outcomes, which are not mutually exclusive:

- Affect the function of DRhoGEF2 e.g. affect activation of Rho1
- Affect the function of the PDZ-binding protein

Put another way, proteins interacting with the PDZ domain could be either upstream or downstream of DRhoGEF2 in a signalling pathway.

If the DRhoGEF2 PDZ domain does not interact with DMCC *in vivo*, it may interact with the other potential targets picked up in the yeast 2-hybrid screen, mec-2 and group 3. A rapid approach to test whether these genes are likely to function genetically with DRhoGEF2 would be to over-express DRhoGEF2 in S2 culture cells and look for rescue of the resulting contraction phenotype by addition of dsRNA to mec-2 or group 3. This would indicate whether the protein products of either of these genes are likely to influence DRhoGEF2 signalling. There may also be an RNAi phenotype associated with mec-2 or group 3 on their own, and if this were the case, addition of DRhoGEF2 dsRNA could be tested for enhancement/suppression of these phenotypes. Another indication that these genes are interacting with DRhoGEF2 *in vivo* would be if their protein products biochemically interacted with full-length DRhoGEF2, and this could be tested via co-immunoprecipitation experiments. Both the biochemical and RNAi tests could also be applied to the other PXTXL genes, or other potential DRhoGEF2 interactors. If positive results were achieved with these tests, the longer term projects of generating a UAS-RNAi transgenic fly, a mutant, or a UAS overexpression fly could be employed to test for genetic interaction *in vivo*.

It is also possible that there are further targets for the PDZ domain of DRhoGEF2 that were not identified in the screen. The microtubule-binding protein, EB1, has been identified as an interactor of DRhoGEF2,

and its site of interaction is unknown (although it does not have a classic PDZ-binding motif at its C-terminus). Other potential interactors could be identified from RNAi screens for enhancement or suppression of the DRhoGEF2 RNAi phenotype, or genetic modifier screens using whole flies. These sort of screens would not be specific for proteins interacting with the PDZ domain, but if any potential interactors from these screens carried PDZ-binding motifs at their C-terminus it could indicate that they interact with DRhoGEF2 via this domain.

It is possible that the PDZ domain of DRhoGEF2 does not have functional significance for DRhoGEF2. A mutant form of DRhoGEF2 that lacked the PDZ domain of DRhoGEF2 could be generated and used to attempt rescue of the DRhoGEF2 null phenotype. If this mutant were to give none or only a partial rescue, this would imply that the PDZ domain does have a functional role, and the resulting phenotype may give some clue as to this role. There is the possibility that deletion of the PDZ domain from the protein would lead to an unstable protein being produced, and therefore wild-type DRhoGEF2 protein levels would need to be confirmed in this mutant.

Another way to test the function of the PDZ domain would be to generate a protein that carried a PDZ-binding motif and over-express this in either *Drosophila* cells in culture or in whole flies by making a transgenic fly, and observe any resulting phenotype. For example, a short polypeptide with TXL at its C-terminus might be expected to interact with the DRhoGEF2 PDZ domain and prevent targets from binding. If DRhoGEF2 requires a protein to bind to its PDZ domain in order for it to be activated, a short polypeptide such as this would be expected to act as a dominant negative. Similarly, if the PDZ targets required binding to DRhoGEF2 in order to become activated themselves, this would also act as a dominant negative for them. However, a TXL motif would be likely to affect other PDZ domains in addition to DRhoGEF2, so a longer sequence such as

PXTXL (if this were determined to be specific for DRhoGEF2) could be employed.

Similarly, the PDZ domain of DRhoGEF2 on its own could be over-expressed in a different dominant negative approach (it may have to be expressed as a fusion since it may not be stable when isolated). In this case targets would be saturated by the isolated PDZ domain and therefore not able to bind DRhoGEF2 and activate it. However, the targets themselves may be hyper-activated since there would be an excess of PDZ domain for them to bind to. It is possible that binding of a target to the PDZ domain of DRhoGEF2 may inhibit, rather than activate, signalling from either the target or DRhoGEF2, and therefore lead to different outcomes for each of these dominant negative approaches. The results of these experiments may be difficult to interpret, as is the case for many dominant negative approaches, but they could certainly give an idea as to whether the PDZ domain does indeed have a function, and some clues as to what that function is.

In conclusion, both MCC and RhoGEF require further study to determine whether the interaction between them is real *in vivo*, and if so, to understand the significance of the interaction. Since studies using *Drosophila* have not proved fruitful in determining function, studies on cell cycle progression and regulation of the actin cytoskeleton in mammalian cells will be required to determine the precise function of these proteins. Of particular interest is how these proteins influence cellular growth, since both MCC and the Rho GTPases are implicated in the progression to cancer.

6.4. Chapter 6 References

- Banerjee, J. and P. B. Wedegaertner (2004). "Identification of a novel sequence in PDZ-RhoGEF that mediates interaction with the actin cytoskeleton." Mol Biol Cell **15**(4): 1760-75.
- Fearon, E. R. and B. Vogelstein (1990). "A genetic model for colorectal tumorigenesis." Cell **61**(5): 759-67.

- Olson, M. F., A. Ashworth, et al. (1995). "An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1." Science **269**(5228): 1270-2.
- Prusiner, S. B. (1998). "Prions." Proc Natl Acad Sci U S A **95**(23): 13363-83.
- Swiercz, J. M., R. Kuner, et al. (2002). "Plexin-B1 directly interacts with PDZ-RhoGEF/LARG to regulate RhoA and growth cone morphology." Neuron **35**(1): 51-63.
- Togashi, H., K. Nagata, et al. (2000). "Functions of a rho-specific guanine nucleotide exchange factor in neurite retraction. Possible role of a proline-rich motif of KIAA0380 in localization." J Biol Chem **275**(38): 29570-8.

Appendix 1: DNA and protein sequences

CG6156 (3002bp)

```
1      CTTCCGCGCA GCTTGAGCTG CCTGTGCTGC AGCGATACGC CATCGGTGCC CAACGACAAG
61     AGACGCAGTT GGGAGTATAC GCTGATGGCT CATCCTCTGG CCAGACGGCC CATAAACGTG
121    ATGACCCCAA GAATAACCAC TTAGAAAGCT CTAATCCGCC TCTTTTCTTT TGATTACCTT
181    CCAGGTTGAG CAACTGTCGG TACAACAGCT AGAAAACCGG GTGCGGGATC TAACGCAGCG
241    TCTGCAGCAG GCGGAAAGAC AGCTCACTGA GAGCAACACA GAGCGGAAA TAGCCACAA
301    GCGCTTGGAG GTTGTCAGCC AGGCCCACGA GTGTCGCATC ACAGAGATGC ACTGTGTCAT
361    CGCCGAGTTA AGCAAGAAAT TGCAGCAGTA GCAGGACCAC GTTATCATGG AGGAGCAGGA
421    GCCCCAGCGC AGCGGTGCGT AGCAGTTTCA AATGAGGGAA ACAAACCCCT AACCGACAGT
481    CTTGTTTTCA GAACTCAGCT TTCAGGAGGG TTCAGTGTAC AATTCCGAGC TTAACCTCAC
541    CAATCCTGAT GCCGAATGTC AAACAGAACC ACTGGAAGAC TTCGAGGGCG CCTGCAGTAC
601    CACCACTGTG GGTAAACGTTG CCCACAAACG GCCGGAACGT AGCCATAAGG GACAAGTGGA
661    GGCAGTGCAG GCGGAAGTTC TGCACTTTGA AGCTCGAATC GCCCTCCTCC AGTCCGAGAT
721    TTCCACCAAG GATGCTGCTG TGGTCGAGGA ACAGACCAA GTCGCCTTCG ACTGCGAATC
781    GGAAGTCAAC GAGTGGCGAC AGCGACTGAA TGATTTGAAT GGTAATATAA ATAATCCTGA
841    TGATAAAAGA ATGGTTACAC ATTTAGATGC TTTCAGTTTG CACTTCCTTG ACAAGTCCGC
901    AAAAACGCAT TCCAGCGGTA CCGAAAATGG CTGAACGGGT TAAGTTGCGG TGCGCCAGCA
961    AACATGAATC CGGAGAAGAT CCATCCCAAG ATACTTCAT T GAGCAACGAG GTGTGTTGAC
1021   CAAAACTTTA TTCAGAGAAT ACATTTCAAC CCATACAACC CATAATTTTA AGCAAATTAA
1081   TCTGGTCGAA CATTTGGTGT CGGAGCTAAA GGAGCAAAAT CTATATATGG AGAATTTAT
1141   GGAGCCCTTA CATTTGAGCA AAGACTTGGA GCGACTGCAA CGACGTGTTG AACAAATTGA
1201   GATGCGAAAC ACCATGCTGG CACTGACGCT AGACGAATGC AAGGAGCACA CTGAGCATCT
1261   GTATCTGCTA TGCGGAAAGT ACGAGTCCAA TGCGGTGCT CTTCAGTTGG CGCTTAATTG
1321   CAGTGATCGC GCCATCGAGG CCTACGACGT AATGTTGGCT CTGCTCGAAA GCAAGTAAAGT
1381   ACAATCACCA TAAGACGCGG GTTGTGTGCC CTATTTTACA ATTCATCACT TCCAGGTTGG
1441   CACTGCTGGG GGAGAAATCA GTGGCAGCGG AAGAGAGTCG ACGATCGGTG GAGGCGGTGG
1501   CCAGGCACCT CTTAGCCCGT TTGGATGCGG AGAAAAACGT TTGTGAGAAC AGCCTGGGAC
1561   CGTGGAACA CAACATCAAC CTGGGCCAG AGGATGCCCC AAAAACTGGC CGCCCGTGGT
1621   GTGCCGACGA CGACAACCGC CTGCGTTACC ACGTCTCCAA GCTGAAGGGA CGTCGTTCCA
1681   ATGTCCAGCA TACCATTGTC AGTTTGGAAT CACCCTTCAG CGACATATAC GAAAGAAAGC
1741   GCCTGGCTTT GGAAGAGGAG CACGAAC TTCGAGCTGCG GGATTCGAAT TTGCAGCTGA
1801   TGGAGACAGC AGTGATTATG CAAGAAATAC TCGAGCTGCG GGATTCGAAT TTGCAGCTGA
1861   AGACAAAAAT GGAAGAGGCC GAGCAGGAAC GGCAGAACGC CAACGAACGA GTGGGCATAC
1921   TCCACGAAGC CTTAAAGCAG CTACAGGCAA ACAACCGGGT CTCGTATTTC GAGGCGGAGC
1981   ATGCGGCTCT CACAGAGCAG CAGTTGGTGG AGGCCTTAAC TCGAGAAACG GAGCTCAAGG
2041   GTCGCATACA GACGCTATTG GCAAATGTAA CAGCTTCGCA GAAGGCCTTC GACGAAAAAT
2101   ACGAGCAACT GCATCAGAAC GTGCGTGAAC TGCAGAAATC CAACCAAGTAA GATTCGTTTTT
2161   ATTTTGACTG TTTTCATACAT AATTTCTTCA TATTTCCCTT TTAGCAATCT GGGCCAAATG
2221   TTAGATCACA CCAAGCGCAA GTACCAGCTG CGGGTGAGGA AGCTAGAGCA GAAGATTGTT
2281   GACCTGCGGC TGGACTACGA GCAAGGCCAT AACCATGTTC CTGAGACTAC TCTGTAGGAA
2341   ACGCATGACG ACCTTCCAGG AGGACTGAGC AACTGGGCGT GGAGCGGGCG CGCTTGTAC
2401   ATATCCCGCA AAAGATCTTT CCCTCTCGAT CATATCGCTT ATTTTACCTT TAATGCAATG
2461   ACCACCCGCC CAGCAGTCGC CGTGGATCCG CCATGACCAA GACAGTATCA CGAATGTGCC
2521   ATGCGACTTA ATTGCAGTCG GCTGGGCTT GGGCGGTGTC CCTGCCTCGC AGATAAGGCT
2581   GCAGTAAC TAATCTCAAA TACATACATA TTCAGAGGTG CCATGAGCAA GGCTCCACCA
2641   ACCTATCCAT AGATCCGTGG GATTGGAGCA TCCGTCTATG GGCCACAAGC AATTACATAT
2701   ACACACATAG GAATAGACAA ACTAAGGAGT TATTCAAGAC GCATACACGG GATCCTATAT
2761   TTATACAATG TATTCGCATT TTGCTTGTTA TATGATTCAA TATGTATTTA AAACGTGACA
2821   AAATATAAAA CGTCTACTAA AACTCGATAA AATTCACGAA TTTGAAATTG GCTCCAATAA
2881   TTAATCGATA AGCTAATGAA ACTAACTTAA GCATCCATCA CCATTTCTGTG TAGTGTGCA
2941   AATCACCAGA ATCAAGTATA TGAGTACTTG TGAAATAAAA ATGTTTTGAA AATTTAACTT
3001   AA
```

cDNAa (2486bp)

```
1   CTTCCGCGCA GCTTGAGCTG CCTGTGCTGC AGCGATACGC CATCGGTGCC
51  CAACGACAAG AGACGCAGTT GGGAGTATAC GCTGATGGCT CATCCTCTGG
101 CCAGACGGCC CATAAACGTT GAGCAACTGT CCGTACAACA GCTAGAAAAC
151 CGGGTGCGGG ATCTAACGCA GCGTCTGCAG CAGGCGGAAA GACAGCTCAC
201 TGAGAGCAAC ACAGAGCGGG AAATATGCCA CAAGCGCTTG GAGGTTGTCA
251 GCCAGGCCCA CGAGTGTGCG ATCACAGAGA TGCAGTGTGT CATCGCCGAG
301 TTAAGCAAGA AATTGCGCAG TAAGCAGGAC CACGTTATCA TGGAGGAGCA
351 GGAGCCCGAC GGAGCGGAAC TCAGCTTTCA GGAGGGTTCA GTGTACAATT
401 CCGAGCTTAA CCTCACCAAT CCTGATGCCG AATGTCAAAC AGAACCCTG
451 GAAGACTTCG AGGGCGCCTG CAGTACCACC AGCGTGGGTA ACGTTGCCCC
501 CAACTGAGCC ATAAGGGACA AGTGGAGGCA CTGCAGGAGG
551 AAGTTCTGCA ATTTGAGAGC CGAATCGCCC TCCTCCAGTC CGAGATTTCC
601 ACCAAGGATG CTGCTGTGGT CGAGGAACAG ACCAAAGTCG CCTTCGACTG
651 CGAATCGGAA GTCAACGAGT GCGGACAGCG ACTGAATGAT TTGAATGTTT
701 GCACTTCCTT GACAAGTCCG CAAAAACGCA TTCCAGCGGT ACCGAAAATG
751 GCTGAACGGG TTAAGTTGCG GTGCGCCAGC AAACATGAAT CCGGAGAAGA
801 TCCATCCCAA GATACTTCAT TGAGCAACGA GCAAATTAAT CTGGTCGAAC
851 ATTTGGTGTC GGAGCTAAAG GAGCAAAATC TATATATGGA GAACTTTATG
901 GAGCCCTCAT ATTTGAGCAA AGACTTGGAG CGACTGCAAC GACGTGTTGA
951 ACAATTGGAG ATGCGAAAACA CCATGCTGGC ACTGACGCTA GACGAATGCA
1001 AGGAGCACAC TGAGCATCTG TATCTGCTAT GCGGAAAGTA CGAGTCCAAT
1051 GCGGTTGCTC TTCAGTTGGC GCTTAATTGC AGTGATCGCG CCATCGAGGC
1101 CTACGACGTA ATGTTGGCTC TGCTCGAAAG CAAGTTGGCA CTGCTGGGGG
1151 AGAAATCAGT GGCAGCGGAA GAGAGTCGAC GATCGGTGGA GCGGTGGCC
1201 AGGCACCTG TAGCCCGTTT GGATAGCGAG AAAAACGTTT GTGAGAACAG
1251 CCTGGGACCG TGGCAACACA ACATCAACCT GGGCCAGAG GATGCCCCAA
1301 AAAGTGGCCG CCCGTGGTGT GCCGACGACG ACAACCGCCT GCGTTACCAC
1351 GTCTCCAAGC TGAAGGGACG TCGTTCCAAT GTCCAGCATA CCATTGTCAG
1401 TTTGGAATCA CCCTTCAGCG ACATATACGA AAGAAAGCGC CTGGCTTTGG
1451 AAAAGGAGCA CGAACTTCGG AGCGCGGACA AGAAGTCACC CATTGACTTG
1501 GAGACAGCAG TGATTATGCA AGAAATACTC GAGCTGCGGG ATTCTGAATT
1551 GCAGCTGAAG AAAAAAATGG AAGAGGCCGA GCAGGAACGG CAGAACGCCA
1601 ACGAACGAGT GGGCATACTC CACGAAGCCC TAAAGCAGCT ACAGGCAAAC
1651 AACC GGGTCTT AACTGTCGGA GGCGGAGCAT GCGGCTCTCA CAGAGCAGCA
1701 GTTGGTGGAG GCCTTAATCT GAGAAACGGA GCTCAAGGGT CGCATACAGA
1751 CGCTATTGGC AAATGTAACA GCTTCGCAGA AGGCCTTCGA CGAAAAATAC
1801 GAGCAACTGC ATCAGAACGT GCGTGAACGT CAGAAATCCA ACCACAATCT
1851 GGGCCAAATG TTAGATCACA CCAAGCGCAA GTACCAGCTG CGGGTGAGGA
1901 AGCTAGAGCA GAAGATTGTT GACCTGCGGC TGGACTACGA GCAAGGCCAT
1951 AACCATTGTT CTGAGACTAC TCTGTAGGAA ACGCATGACG ACCTTCCAGG
2001 AGGACTGAGC AACTGGGCGT GGAGCGGGCG CGCTTGTCAC ATATCCCGCA
2051 AAAGATCTTT CCCTCTCGAT CATATCGCTT ATTTTACCTT TAATGCAATG
2101 ACCACCCGCC CAGCAGTCGC CGTGGATCCG CCATGACCAA GACAGTATCA
2151 CGAATGTGCC ATGCGACTTA ATTGCAGTCG GCTGGGGCTT GGGCGGTGTC
2201 CCTGCCTCGC AGATAAGGCT GCAGTAACCTA CATACTCAA TACATACATA
2251 TTCAGAGGTG CCATGAGCAA GGCTCCACCA ACCTATCCAT AGATCCGTGG
2301 GATTGGAGCA TCCGTCTATG GGCCACAAGC AATTACATAT ACACACATAC
2351 GAATAGACAA ACTAAGGAGT TATTCAAGAC GCATACACGG GATCCTATAT
2401 TTATACAATG TATTCGCATT TTGCTTGTTA TATGATTCAA TATGTATTTA
2451 AAAGTGTACA AAATATAAAA CGTCTACTAA AACTCG
```

cDNAb (2702bp)

```
1   CTTCCGCGCA GCTTGAGCTG CCTGTGCTGC AGCGATACGC CATCGGTGCC
51  CAACGACAAG AGACGCAGTT GGGAGTATAC GCTGATGGCT CATCCTCTGG
101 CCAGACGGCC CATAAACGTT ATGACCCCAA GAATAACCAC TTAGAAAGCT
151 CTAATCCGCC TCTTTTCTTT TGATTCACCT CCAGGTTGAG CAACTGTCCG
201 TACAACAGCT AGAAAACCGG GTGCGGGATC TAACGCAGCG TCTGCAGCAG
```

```

251 GCGGAAAGAC AGCTCACTGA GAGCAACACA GAGCGGGAAA TATGCCACAA
301 GCGCTTGGAG GTTGTGAGCC AGGCCACGA GTGTGCGATC ACAGAGATGC
351 ACTGTGTCAT CGCCGAGTTA AGCAAGAAAT TGCGCAGTAA GCAGGACCAC
401 GTTATCATGG AGGAGCAGGA GCGGACGGC AGCGAACTCA GCTTTCAGGA
451 GGGTTCAGTG TACAATTCCG AGCTTAACCT CACCAATCCT GATGCCGAAT
501 GTCAAACAGA ACCACTGGAA GACTTCGAGG GCGCCTGCAG TACCACCAGC
551 GTGGGTAACG TTGCCCACAA ACCGCCGAA CTGAGCCATA AGGGACAAGT
601 GGAGGCACTG CAGGAGGAAG TTCTGCACTT GAGAGCTCGA ATCGCCCTCC
651 TCCAGTCCGA GATTTCACC AAGGATGCTG CTGTGGTCGA GGAACAGACC
701 AAAGTCGCCT TCGACTGCGA ATCGGAAGTC AACGAGTGCG GACAGCGACT
751 GAATGATTTG AATGTTTGCA CTTCCTTGAC AAGTCCGCAA AAACGCATTC
801 CAGCGGTACC GAAAATGGCT GAACGGGTTA AGTTGCGGTG CGCCAGCAAA
851 GTTGGCACTG CTGGGGGAGA AATCAGTGGC AGCGGAAGAG AGTCGACGAT
901 AATTAATCTG GTCGAACATT TGGTGTGCGA GCTAAAGGAG CAAAATCTAT
951 ATATGGAGAA CTTTATGGAG CCCCTACATT TGAGCAAAGA CTTGGAGCGA
1001 CTGCAACGAC GTGTGAACA ATTGGAGATG CGAAACACCA TGCTGGCACT
1051 GACGCTAGAC GAATGCAAGG AGCACACTGA GCATCTGTAT CTGCTATGCG
1101 GAAAGTACGA GTCCAATGCG GTTGCTCTTC AGTTGGCGCT TAATTGCAGT
1151 GATCGCGCCA TCGAGGCCTA CGACGTAATG TTGGCTCTGC TCGAAAGCAA
1201 GTTGGCACTG CTGGGGGAGA AATCAGTGGC AGCGGAAGAG AGTCGACGAT
1251 CGGTGGAGGC GGTGGCCAGG CACCTGCTAG CCCGTTTGGA TAGCGAGAAA
1301 AACGTTTGTG AGAACAGCCT GGGACCGTGG CAACACAACA TCAACCTGGG
1351 CCCAGAGGAT GCCCAAAAA CTGGCCGCCC GTGGTGTGCC GACGACGACA
1401 ACCGCCTGCG TTACCACGTC TCCAAGCTGA TGGGACGTCG TTCCAATGTC
1451 CAGCATACCA TTGTCAGTTT GGAATCACCC TTCAGCGACA TATACGAAAG
1501 AAAGCGCCTG GCTTTGAAA AGGAGCACGA ACTTCGGAGC GCGGACAAGA
1551 AGTCACCCAT TGACTTGGAG ACAGCAGTGA TTATGCAAGA AATACTCGAG
1601 CTGCGGGATT CGAATTTGCA GCTGAAGACA AAAATGGAAG AGGCCGAGCA
1651 GGAACGGCAG AACGCCAACG AACGAGTGGG CATACTCCAC GAAGCCCTAA
1701 AGCAGCTACA GGCAACAAC CGGGTCTCGT ATTCGGAGGC GGAGCATGCG
1751 GCTCTCACAG AGCAGCAGTT GGTGGAGGCC TTAACCTCGAG AAACGGAGCT
1801 CAAGGGTCGC ATACAGACGC TATTGGCAAA TGTAACAGCT TCGCAGAAGG
1851 CCTTCGACGA AAAATACGAG CAACTGCATC AGAACGTGCG TGAAGTGCAG
1901 AAATCCCAAC ACAATCTGGG CCAAATGTTA GATCACACCA AGCGCAAGTA
1951 CCAGCTGCGG GTGAGGAAGC TAGAGCAGAA GATTGTTGAC CTGCGGCTGG
2001 ACTACGAGCA AGGCCATAAC CATGTTCCCTG AGACTACTCT GTAGGAAACG
2051 CATGACGACC TTCCAGGAGG ACTGAGCAAC TGGGCGTGGA GCGGGCGCGC
2101 TTGTCACATA TCCCGCAAAA GATCTTTCCC TCTCGATCAT ATCGCTTATT
2151 TTACCTTTAA TGCAATGACC ACCCGCCCAG CAGTCGCCGT GGATCCGCCA
2201 TGACCAAGAC AGTATCACGA ATGTGCCATG CGACTTAATT GCAGTCGGCT
2251 GGGGCTTGGG CGGTGTCCCT GCCTCGCAGA TAAGGCTGCA GTAAGTACAT
2301 ACTCAAATAC ATACATATTC AGAGGTGCCA TGAGCAAGGC TCCACCAACC
2351 TATCCATAGA TCCGTGGGAT TGGAGCATCC GTCTATGGGC CACAAGCAAT
2401 TACATATACA CACATACGAA TAGACAAACT AAGGAGTTAT TCAAGACGCA
2451 TACACGGGAT CCTATATTTA TACAATGTAT TCGCATTTTG CTTGTTATAT
2501 GATTCAATAT GTATTTAAAA CTGTACAAAA TATAAAACGT CTACTAAAAC
2551 TCGATAAAAT TCACGAATTT GAAATTGGCT CCAATAATTA ATCGATAAGC
2601 TAATGAAACT AACTTAAGCA TCCATCACCA TTTCGTGTAG TGTGCAAAAT
2651 CACCAGAATC AAGTATATGA GTACTTGTGA AATAAAAATG TTTTGAAAAAT
2701 TT

```

cDNAc (2834bp)

```

1 GGAGTAATTT CGAGTTTTTG AATTAAAAAT GCCTTAAGAA AAGTTTTTTAA
51 TGAGCCATGT AGAGTTAACA CCAGTGCCAC ATATCCAGGG AATGGAGCAC
101 CAATCAAAAAG TCATCATCAA GCGGCCCAA TCAGGCCAAT TCTAGTCGGA
151 AAGTCGTTGC AGCCGCCCAT TGCCACCATG TCGAATGATG TGCAGGTCGC
201 CCGGGTGGCC AAGATAGCTA CCGATGTGCC GCGTCGCAGT GGCAAGCAGC
251 GTGACTCCAG CGGATTCAG GGCAAGCACT CCGGCAGCGC GGCTGGCGAG
301 GATTTTCGAGT ACGTCTTCGG TAGCATTTTCG CCGCGCGGTG GAGGACCCGG

```

```

351 TGGCAGGCAC TTGGTGGGAT CCTCCGACCT GGACTCTCCG GAGCACACGC
401 AGCGGGACAC CACCGAGAGT GACAACAACA TATCCAGCTG CTCCACGCTA
451 GACATTGTCA ACAAAGTTGA GCAACTGTCT GTACAACAGC TAGAAAACCG
501 GGTGCGGGAT CTAACGCAGC GTCTGCAGCA GGCGGAAAGA CAGCTCACTG
551 AGAGCAACAC AGAGCGGGAA ATATGCCACA AGCGCTTGGA GGTGTGCAGC
601 CAGGCCCCAG AGTGTGCGAT CACAGAGATG CACTGTGTCA TCGCCGAGTT
651 AAGCAAGAAA TTGCGCAGTA AGCAGGACCA CGTTATCATG GAGGAGCAGG
701 AGCCCGACGG CAGCGAACTC AGCTTTCAGG AGGGTTCAGT GTACAATTCC
751 GAGCTTAACC TCACCAATCC TGATGCCGAA TGTCAAACAG AACCCTGGA
801 AGACTTCGAG GCGCCTGCA GTACCACCAG CGTGGGTAAC GTTGCCACA
851 AACC GCCGA ACTGAGCCAT AAGGGACAAG TGGAGGCACT GCAGGAGGAA
901 GTTCTGCACT TGAGAGCTCG AATCGCCCTC CTCCAGTCCG AGATTTCAC
951 CAAGGATGCT GCTGTGGTCG AGGAACAGAC CAAAGTCGCC TTCGACTGCG
1001 AATCGGAAGT CAACGAGTGC GGACAGCGAC TGAATGATTT GAATGTTTGC
1051 ACTTCCTTGA CAAGTCCGCA AAAACGCATT CCAGCGGTAC CGAAAATGGC
1101 TGAACGGGTT AAGTTGCGGT GCGCCAGCAA ACATGAATCC GGAGAAGATC
1151 CATCCCAAGA TACTTCATTG AGCAACGAGC AAATTAATCT GGTGCAACAT
1201 TTGGTGTCTG AGCTAAAGGA GCAAAATCTA TATATGGAGA ACTTTATGGA
1251 GCCCTACAT TTGAGCAAAG ACTTGGAGCG ACTGCAACGA CGTGTGTAAC
1301 AATTGGAGAT GCGAAACACC ATGCTGGCAC TGACGCTAGA CGAATGCAAG
1351 GAGCACACTG AGCATCTGTA TCTGCTATGC GGAAAGTACG AGTCCAATGC
1401 GGTGCTCTT CAGTTGGCGC TTAATTGCAG TGATCGCGCC ATCGAGGCCCT
1451 ACGACGTAAT GTTGGCTCTG CTCGAAAGCA AGTTGGCACT GCTGGGGGAG
1501 AAATCAGTGG CAGCGGAAGA GAGTCGACGA TCGGTGGAGG CGGTGGCCAG
1551 GCACCTGCTA GCCCGTTTGG ATAGCGAGAA AAACGTTTGT GAGAACAGCC
1601 TGGGACCGTG GCAACACAAC ATCAACCTGG GCCCAGAGGA TGCCCCAAAA
1651 ACTGGCCGCC CGTGGTGTGC CGACGACGAC AACC GCCTGC GTTACCACGT
1701 CTCCAAGCTG AAGGGACGTC GTTCCAATGT CCAGCATACC ATTGTCAGTT
1751 TGAATCACC CTTCAGCGAC ATATACGAAA GAAAGCGCCT GGCTTTGGAA
1801 AAGGAGCACG AACTTCGGAG CGCGGACAAG AAGTCACCCA TTGACTTGGA
1851 GACAGCAGTG ATTATGCAAG AAATACTCGA GCTGCGGGAT TCGAATTTCG
1901 AGCTGAAGCA AAAAAATGGA GAGGCCGAGC AGGAACGGCA GAACGCCAAC
1951 GAACGAGTGG GCATACTCCA CGAAGCCCTA AAGCAGCTAC AGGCAAACAA
2001 CCGGGTCTCG TATTCGGAGG CGGAGCATGC GGCTCTCACA GAGCAGCAGT
2051 TGGTGGAGGC CTTAACTCGA GAAACGGAGC TCAAGGGTCG CATAAGACG
2101 CTATTGGCAA ATGTAACAGC TTTCGAGAAG GCCTTCGACG AAAAAATACGA
2151 GCAACTGCAT CAGAACGTGC GTGAAC TGCA GAAATCCAAC CACAATCTGG
2201 GCCAAATGTT AGATCACACC AAGCGCAAGT ACCAGCTGCG GGTGAGGAAG
2251 CTAGAGCAGA AGATTGTTGA CCTGCGGCTG GACTACGAGC AAGGCCATAA
2301 CCATGTTTCT GAGACTACTC TGTAGGAAAC GCATGACGAC CTTCCAGGAG
2351 GACTGAGCAA CTGGGCGTGG AGCGGGCGCG CTGTGCACAT ATCCCGCAAA
2401 AGATCTTTCC CTCTCGATCA TATCGCTTAT TTTACCTTTA ATGCAATGAC
2451 CACCCGCCCA GCAGTCGCCG TGGATCCGCC ATGACCAAGA CAGTATCACG
2501 AATGTGCCAT GCGACTTAAT TGCAGTCGGC TGGGGCTTGG GCGGTGTCCC
2551 TGCCTCGCAG ATAAGGCTGC AGTAACTACA TACTCAAATA CATAATATT
2601 CAGAGGTGCC ATGAGCAAGG CTCCACCAAC CTATCCATAG ATCCGTGGGA
2651 TTGGAGCATC CGTCTATGGG CCACAAGCAA TTACATATAC ACACATACGA
2701 ATAGACAAAC TAAGGAGTTA TTCAAGACGC ATACACGGGA TCCTATATTT
2751 ATACAATGTA TTCGCATTTT GCTTGTTATA TGATTCAATA TGTATTTAAA
2801 ACTGTACAAA ATATAAAACG TCTACTAAAA CTCG

```

cDNA_{Ad} (1249bp)

```

1 GGAGTAATTT CGAGTTTTTG AATTAAAAAT GCCTTAAGAA AAGTTTTTAA
51 TGAGCCATGT AGAGTTAACA CCAGTGCCAC ATATCCAGGG AATGGAGCAC
101 CAATCAAAAG TCATCATCAA GCGGCCCAAA TCAGGCCAAT TCTAGTCGGA
151 AAGTCGTTTG AGCCGCCCAT TGCCACCATG TCGAATGATG TGCAGTCTCG
201 CCGGGTGGCC AAGATAGCTA CCGATGTGCC GCGTCGCAGT GGCAAGCAGC
251 GTGACTCCAG CGGATTCAGG GGCAAGCACT CCGGCAGCGC GGCTGGCGAG
301 GATTTCGAGT ACGTCTTCGG TAGCATTTTC CCGCGCGGTG GAGGACCCGG

```

351	TGGCAGGCAC	TTGGTGGGAT	CCTCCGACCT	GGACTCTCCG	GAGCACACGC
401	AGCGGGACAC	CACCGAGAGT	GACAACAACA	TATCCAGCTG	CTCCACGCTA
451	GACATTGTCA	ACAAAGTTGA	GCAACTGTCT	GTACAACAGC	TAGAAAAACCG
501	GGTGCGGGAT	CTAACGCAGC	GTCTGCAGCA	GGCGGAAAGA	CAGCTCACTG
551	AGAGCAACAC	AGAGCGGGAA	ATATGCCACA	AGCGCTTGGA	GGTTGTCAGC
601	CAGGCCCACG	AGTGTCGCAT	CACAGAGATG	CACTGTGTCA	TCGCCGAGTT
651	AAGCAAGAAA	TTGCGCAGTA	AGCAGGACCA	CGTTATCATG	GAGGAGCAGG
701	AGCCCGACGG	CAGCGAACTC	AGCTTTCAGG	AGGGTTCAGT	GTACAATTCC
751	GAGCTTAACC	TCACCAATCC	TGATGCCGAA	TGTCAAACAG	AACCACTGGA
801	AGACTTCGAG	GGCGCCTGCA	GTACCACCAG	CGTGGGTAAC	GTTGCCCCACA
851	AACCGCCGGA	ACTGAGCCAT	AAGGGACAAG	TGGAGGCACT	GCAGGAGGAA
901	GTTCTGCAC	TGAGAGCTCG	AATCGCCCTC	CTCCAGTCCG	AGATTTCAC
951	CAAGGATGCT	GCTGTGGTCT	AGGAACAGAC	CAAAGTCGCC	TTCGACTGCG
1001	AATCGGAAGT	CAACGAGTGC	GGACAGCGAC	TGAATGATTT	GAATGTTTGC
1051	ACTTCCTTGA	CAAGTCCGCA	AAAACGCATT	CCAGCGGTAC	CGAAAATGGC
1101	TGAACGGGTT	AAGTTGCGGT	GCGCCAGCAA	ACATGAATCC	GGAGAAGATC
1151	CATCCCAAGA	TACTTCATTG	AGCAACGAGG	GTGTGTTGAC	CAAACTTTA
1201	TTCAGAGAAT	ACATTTCAAC	CCATACAACC	CAATCGAATT	CCCGCGGCC

HsMCC cDNA (4180bp)

1	CCTCCTGCAG	CAATGGCTCG	TCCGTGAAAC	GCGAGCCACG	GCTGCTCTTT	TTAAGAGTGC
61	CTGCATCCTC	CGTTTGCGCT	TCGCAACTGT	CCTGGGTGAA	AATGGCTGTC	TAGACTAAAA
121	TGTGGCAGAA	GGGACCAAGC	AGTGGATATT	GAGCCTGTGA	AGTCCAACTC	TTAAGCTCCG
181	AGACCTGGGG	GACTGAGAGC	CCAGCTCTGA	AAAGTGCATC	ATGAATTCCG	GAGTTGCCAT
241	GAAATATGGA	AACGACTCCT	CGGCCGAGCT	GAGTGAGCTC	CATTGAGCAG	CCCTGGCATC
301	ACTAAAGGGA	GATATAGTGG	AACTTAATAA	ACGTCTCCAG	CAAACAGAGA	GGGAACGGGA
361	CCTTCTGGAA	AAGAAATTGG	CCAAGGCACA	GTGCGAGCAG	TCCCACCTCA	TGAGAGAGCA
421	TGAGGATGTC	CAGGAGCGAA	CGACACTTCG	CTATGAGGAA	CGCATCACAG	AGCTCCACAG
481	CGTCATTGCG	GAGCTCAACA	AGAAGATAGA	CCGTCTGCAA	GGCACCACCA	TCAGGGAGGA
541	AGATGAGTAC	TCAGAACTGC	GATCAGAACT	CAGCCAGAGC	CAACACGAGG	TCAACGAGGA
601	CTCTCGAAGC	ATGGACCAAG	ACCAGACCTC	TGTCTCTATC	CCCGAAAACC	AGTCTACCAT
661	GGTTACTGCT	GACATGGACA	ACTGCAGTGA	CCTGAACTCA	GAAGTGCAGA	GGGTGCTGAC
721	AGGGCTGGAG	AATGTTGTCT	GCGGCAGGAA	GAAGAGCAGC	TGCAGCCTCT	CCGTGGCCGA
781	GGTGGACAAG	CACATTGAGC	AGCTCACCAC	AGCCAGCGAG	CACTGTGACC	TGGCTATTAA
841	GACAGTCGAG	GAGATTGAGG	GGGTGCTTGG	CCGGGACCTG	TATCCCAACC	TGGCTGAAGA
901	GAGGTCTCGG	TGGGAGAAGG	AGCTGGCTGG	GCTGAGGGAA	GAGAAATGAG	GCCTGACTGC
961	CATGCTGTGC	AGCAAAGAGG	AAGAAGTAAA	CCGGACTAAG	GCCACCATGA	ATGCCATCCG
1021	GGAAGAGCGG	GACCGGCTCC	GGAGGAGGGT	CAGAGAGCTT	CAAACCTGAC	TACAGAGCGT
1081	GCAGGCCACA	GGTCCCTCCA	GCCCTGGCCG	CCTCACTTCC	ACCAACCGCC	CGATTAACCC
1141	CAGCACTGGG	GAGCTGAGCA	CAAGCAGCAG	CAGCAATGAC	ATTCCCATCG	CCAAGATTGC
1201	TGAGAGGGTG	AAGCTATCAA	AGACAAGGTC	CGAATCGTCA	TCATCTGATC	GGCCAGTCCT
1261	GGGCTCAGAA	ATCAGTAGCA	TAGGGGTATC	CAGCAGTGTG	GCTGAACACC	TGGCCCCTCT
1321	ACTTTCAGAC	TGCTCCAATA	TCCAAGAGAT	TTTCCAAACA	CTCTACTCAC	ACGGATCTGC
1381	CATCTCAGAA	AGCAAGATTA	GAGAGTTTGA	GGTGGAAACA	GAACGGCTGA	ATAGCCGGAT
1441	TGAGCACCTC	AAATCCCAAA	ATGACCTCCT	GACCATAACC	TTGGAGGAAT	GTAAGGCAA
1501	TGCCGAGAGG	ATGAGCATGC	TGGTGGGAAA	ATACGAATCC	AATGCCACAG	CGCTGAGGCT
1561	GGCCTTGAGC	TACAGCGAGC	AGTGCATCGA	AGCCTACGAA	CTCCTCCTGG	CGCTGGCAGA
1621	GAGTGAGCAG	AGCCTCATCC	TGGGGCAGTT	CCGAGCGGCG	GGCGTGGGGT	CCTCCCCCTGG
1681	AGACCAGTCG	GGGGATGAAA	ACATCACTCA	GATGCTCAAG	CGAGCTCATG	ACTGCCGGAA
1741	GACAGCTGAG	AACGCTGCCA	AGGCCCTGCT	CATGAAGCTG	GACGGCAGCT	GTGGGGGAGC
1801	CTTTGCCGTG	AGCGCTGCA	CCGTGCGACC	CTGGGAGAGC	CTTTCCTCCA	ACAGCCACAC
1861	CAGCACAACC	AGCTCCACAG	CCAGTAGTTG	CGACACCGAG	TTCACTAAAG	AAGACGAGCA
1921	GAGGCTGAAG	GATTATATCC	AGCAGCTCAA	GAATGACAGG	GCTGCGGTCA	AGCTGACCAT
1981	GCTGGAGCTG	GAAAGCATCC	ACATCGATCC	TCTCAGCTAT	GACGTCAAGC	CTCGGGGAGA
2041	CAGCCAGAGG	CTGGATCTGG	AAAACGCAGT	GCTTATGCAG	GAGCTCATGG	CCATGAAGGA
2101	GGAGATGGCC	GAGTTGAAGG	CCCAGCTCTA	CCTACTGGAG	AAAGAGAAGA	AGGCCCTGGA
2161	GCTGAAGCTG	AGCACGCGGG	AGGCCAGGAG	GCAGGCCTAC	CTGGTGCACA	TTGAGCACCT
2221	GAAGTCCGAG	GTGGAGGAGC	AGAAGGAGCA	GCGGATGCGA	TCCCTCAGCT	CCACCAGCAG
2281	CGGCAGCAAA	GACAAACCTG	GCAAGGAGTG	TGCTGATGCT	GCCTCCCCAG	CTCTGTCCCT

2341	AGCCGAAC	TC	AGGACAAC	GT	GCAGCGAG	AA	TGAGCTGG	CT	GCGGAGTT	CA	CCAACGCC	AT
2401	TCGTCGAG	AA	AAGAAGTT	GA	AGGCCAGAG	T	TCAAGAGCT	G	GTGAGTGC	CT	TGGAGAGAC	T
2461	CACCAAGAG	C	AGTGAAAT	CC	GACATCAG	CA	ATCTGCAG	AG	TTCGTGAAT	G	ATCTAAAG	CG
2521	GGCCAACAG	C	AACCTGGT	GG	CTGCCATG	A	GAAAGCAA	AG	AAAAAGCA	TC	AAAACAACT	
2581	GAAGAAGTT	A	GAGTCGCA	G	TGATGGCC	AT	GGTGGAGAG	A	CATGAGACC	C	AAGTGAGGA	T
2641	GCTCAAGCA	A	AGAATAGCT	C	TGCTAGAG	G	GGAGAACT	CC	AGGCCACAC	A	CCAATGAA	AC
2701	TTCGCTTTA	A	TCAGCACT	CA	CGCACC	GG	TTCTGCCC	AT	GGGAAGTAA	A	CTGCAGCAG	G
2761	CCACTGGGG	A	CAGAAGGG	CC	CATGTACT	TT	TTGGGAGG	AG	GAGGAAAGG	G	AAGGCTGG	CA
2821	GGTAGGTCG	G	CACTTGGAC	A	ATGGAGTG	CC	CCAACTCA	AC	CCTTGGGG	CG	ACTGGCCAT	G
2881	GTGACATTG	T	GGACTGTAT	C	CAGAGGTG	CC	CGCTCTTC	CC	TCCTGGGG	CC	ACAACAGCG	T
2941	GTAAACACAT		GTTCTGTGC	C	TGCTCAGAG		AGCCTCGTT	T	CTGCTTTCAG		CACCTACT	T
3001	CCCCCTCCT	C	TTCTGGTCT	G	GCGGCTGTG	C	ATCAGTGGG	A	TCCCAGACAT		TTGTTTCTGT	
3061	AAGATTTTCC		ATTGTATCCT		CTTTTTGGTA		GATGCTGGG	C	TCATCTTCTA		GAATCTCGTT	
3121	TCTCCTCTTT		CTCCTGCTT		CATGGGAAA	A	CAGACCTGT	G	TGTGCCTCCA		GCATTTAAAA	
3181	GGACTGCTGA		TTTGTTTACT		ACAGCAAGGC		TTTGGTTTCC		AAGTCCC	GGG	TCTCAACTTT	
3241	AAGATAGAGG		CGGCCATAAG		AGGTGATCTC		TGGGAGTTAT		AGGTCATGGG		AAGAGCGTAG	
3301	ACAGGTGTTA		CTTACAGTCC		CAGATACACT		AAAGTTACAA		ACAGACCACC		ACCAGGACTG	
3361	TGCCTGAACA		ATTTTGTATT		GAGAGAAATA		AAACTTCCTT		CAATCTTCAT		TTTGGAGGCA	
3421	GGGCTGGGAA		GGGAGCGCTC		TCTTGATTCT		GGGATTTCTC		CCTCTCAGTG		GAGCCTTATT	
3481	AATATCCAAG		ACTTAGAGCT		GGGAATCTTT		TTGATACCTG		TAGTGGA	ACT	AAAATTC	TGT
3541	CAGGGGTTTC		TTCAAGAGCT		GAGAAACATT		ATTAGCACTT		CCCGCCCCAG		GGCACTACAT	
3601	AATTGCTGTT		CTGCTGAATC		AAATCTCTTC		CACATGGGTG		CATTTGTAGC		TCTGGACCTG	
3661	TCTCTACCTA		AGGACAAGAC		ACTGAGGAGA		TACTGAACAT		TTTGCAAAAC		TTATCACGCC	
3721	TACTTAAGAG		TGCTGTGTAA		CCCCCAGTTC		AAGACTTAGC		TCCTGTTGTC		ATGACGGGGA	
3781	CAGAGTGAGG		GAATGGTAGT		TAAGGCTTCT		TTTTTGCCCC		CAGATACATG		GTGATGGTTA	
3841	GCATATGGTG		CTTAAAAGGT		TAAATTTCAA		GCAAAATGCT		TACAGGGCTA		GGCAGTACCA	
3901	AAGTAACTGA		ATTATTTTCTAG		GAAGGTCTTC		AATCTTAAAA		CAAATTCATT		ATTCTTTTTC	
3961	AGTTTTACCT		CTTCTCTCTC		AGTTCTACAC		TGATACACTT		GAAGGACCAT		TTACTGTTTT	
4021	TTTCTGTAGC		ACCAGAGAAT		CCATCCAAAG		TTCCCTATGA		AAAAATGTGTT		CCATTGCCAT	
4081	AGCTGACTAC		AAATTAAAGT		TGAGGAGGTT		TCTGCATAGA		GTCTTTATGT		CCATAAGCTA	
4141	CGGGTAGGTC		TATTTTCAGA		GCATGATACA		AATTCCACAG					

DRho GEF2 cDNA - RF isoform (8442bp)

1	AAAAGAAAAG	AACGTCGTGT	GCGTGTGTGAT	GGCGCGATAA	GTGAAAAAAA	TAAAAACATT
61	TTTCGTGTCT	TTATGGAATA	CAATACGAGT	GCGTTTTAAG	TTTAAGTGAA	AAGTAGCAGT
121	GCAAAATATA	TATACACATA	TCAAAACCCA	TTGGAATAGT	GCAACCAAAA	AATCATAAAA
181	CAGAACTAAA	GCAACGAAAA	TCGAGGCAAT	TAAATCACTT	CTGGATTCAT	TTCGGCCCCG
241	TTTTACTAGC	TAATTTAAAA	ATAGCTGCCG	AAAGCATTCC	GCAACCACTT	CTGGTGACTG
301	GCCCCACACC	AATGCGAACA	ATACTCCACG	GGATCCTCCG	ACTTCGAGTG	GCTTACGCAT
361	GCCTTTGAAT	TTGCATCTTC	GCACCCGCAC	AGCAGCCAGC	AAAAGCCATC	GCAGGAAACC
421	CTATGGATGA	CCCATCAATC	AAAAAACGGT	TACTAGATTT	ATATACTGAC	GAACATGAAT
481	ACGATGAGGT	TCAAGAGATA	CCAGAAGAGT	CAAGCATTCA	GCCCCCGGAG	ACATCAACGA
541	GCCATACCAG	CACAAACGGA	TCCAGCCACT	CAGGCCCTGG	AAC TGCAACT	GGACCAGGAG
601	CGACATCTGC	AGGCCCCGTC	GCGGGTGCGC	CACAGTCACC	AGTAATTGTT	GTGGACTCGG
661	TTCCCGAGCT	GCCAGCGCCC	AAGCAGAAAT	CTGTTAAGAA	CTCGAAGAGC	AAACAGAAGC
721	AAAAGCAGTT	GGCGAACAAA	TCGAAGATTTC	CACGATCTCC	TTCGTTGGCG	AGCAGCCTGA
781	GTAGTCTGGC	CAGCAGCCTT	AGTGGTCATA	GGGATCGGGA	CAAAGATCGG	GACAAGGATC
841	GGGAGAACCA	GAACGCCGTG	CCGCCGCAGA	CGCCGCCGTT	GCCACCGAGC	TACAAGCAGA
901	ACCAAATGAA	TGGTGACTCT	ACGGCTGCAG	CTGGTGGTGG	TGTTTCTGCC	CCAGCCACGC
961	CCACCACCGC	CAATAATAAC	AATGCGAGCC	ACAACAACGG	CAGCATAATG	GGCGGAGGCG
1021	TGCAATTGAA	TCAATCGGAC	AACTCCAACC	CCGTTCTCCA	GGCGCCGGGG	GAGCGCAGTA
1081	GCCTCAATCT	GACTCCCCTC	TCCCGGGATC	TGAGTGGTGG	CCACACCCAG	GAGTCCACGA
1141	CGCCAGCGAC	CACGCCGAGT	ACGCCAAGCC	TAGCATTACC	AAAAAATTTT	CAGTATTTAA
1201	CCCTGACTGT	GCGAAAAGAT	AGCAACGGAT	ACGGGATGAA	GGTTTCCGGA	GATAATCCTG
1261	TGTTTGTGGA	GAGCGTTAAA	CCCGGAGGCG	CAGCGGAGAT	TGCAGGCCTG	GTTGCTGGCG
1321	ATATGATACT	AAGGGTAAAC	GGCCATTGAAG	TGCGACTAGA	GAAGCATCCA	ACTGTAGTGG
1381	GTCTTATAAA	AGCCTCGACA	ACTGTTGAGC	TGGCGGTGAA	GCGAAGTCAG	AAGCTAACGC
1441	GACCCTCTTC	GGTATCAGTG	GTAACGCCCT	CGACACCCAT	TCTCTCTGGA	CGAGATCGTA
1501	CCGCTTCAAT	AACTGGGCCA	CAGCCGGTGG	ACAGTATTAA	ACGAAGGGAG	ATGGAGACTT

1561	ACAAAATTCA	GACGTTGCAG	AAAATGCTGG	AGCAGGAGAA	ACTAAATCTG	GAGCGATTAA
1621	AAAGCGATCA	AAATAATCCG	AGCTACAAGC	TATCTGAGGC	GAATATCCGT	AAGCTGCGCG
1681	AGCAACTTCA	TCAAGTGGGA	GCTGAGGATG	CACCGACTGT	TAAACTTCAG	GCGGCCGCTG
1741	GCAATAAGAA	CACAGCATTA	CTAACGCCCA	ACCAAATCCA	ACACTTGTCC	GCGTCCGCCA
1801	CTCACAGCAA	TCAGCAATTC	CATCATCTTC	ACCACCACCA	CAATCTCCAC	AACAACAATT
1861	ATCCGCCACA	GCAACAGCCA	GCTAGCACCT	CACCCGCATT	CCTGTCCCTC	CTGCCGCGTT
1921	CCCTTTTCGTC	CTTGTGCGTG	GGCACGCGCA	AAAACAAGAC	CGAAAAGGAC	CTAACGACTT
1981	CATCTCCCTT	TGGCCTAACG	ACGGATTTC	TGCAGCAGCA	ACGGATGAGC	CACCAAGCGG
2041	AGTCGATGTC	TCAATCGATG	CATCAGCACA	CCAGCACTCC	GACCTCGCAG	CAGTTCTTCC
2101	ATCCGCATCA	GCAACAACAT	CGCTTTAAGG	AAACTGGACC	GACGTCGAAA	GGCAAGAACA
2161	AGTTCCTAAT	TTCGAGGAGT	TTGATCGAGG	AGGATGTGCC	GCCACCACTG	CCGCAGAGGA
2221	ATCCGCCCAG	GCAGTTAAAT	TTGGACCTGA	AGAACGGAAA	CGCGTCGCCG	GGTGGGTAC
2281	ATCTAGTGGC	TCCAGTTTCC	GATCTGGATC	GCGCCACAAG	TCCCCAATTA	AATAGATCCC
2341	AACAACAGCA	ACTGCCTAGC	AGCACTGACA	ACAGCCCGAG	CAATGCCAAG	TCCAAGCGTT
2401	CCAAGATCAA	AACGAAGGCC	CTATCGGATC	CTAAGATGTC	CACCTCAGATG	TTCTTGCAAA
2461	TGGAATCGGC	GAGTGCAGCT	GGAGCAGCCG	GAGGTTCCAT	CGAGGTAGAT	GGGGGTCCAC
2521	CACCGCTACC	GCCTCGCTTG	CCTGGCATGA	TGACGGAGGA	TATGAGCCCG	GGCAGCTGTC
2581	AGAATCTCGC	TCAACCCAAT	TCCGTGGTGA	CTGCCTTCAA	CTATCCTCTG	GTGTCCACAA
2641	CAACAGCGGT	GCAAGACGAT	AACCTGAACA	TTGCCTTTCC	TTTGTCCCAA	CGACCCAACA
2701	TTGTCCAGCA	GCTGCAGCAA	TATCAACAGC	AGCAGCAGCA	TCAAATGAGC	GGTGGCCAGG
2761	CCACAGGCGC	TCTGGGACAG	ACTCCGAATT	TGGGAAAAAA	CAAGCATCGA	CGCGTTGGTT
2821	CTTCACCGGA	CAATATGCAT	CCACGTCATC	CAGATCGGAT	AACGAAAACG	ACTTCGGGCT
2881	CGTGGGAGAT	TGTTGAAAAG	GATGGCGAAT	CCTCCCCGCC	CGGAACACCG	CCGCCTCCAT
2941	ATCTATCCAG	CTCCACATG	ACCGTGCTGG	AAGATCCGAA	TGAGAACAAT	CGTGGAGCAG
3001	CAGCAGCCGG	ACCTGGAGTC	TTTATCGAGT	CGCATCAGTT	TACGCCGATG	CGGGGAGCCT
3061	CTTCTCCGAT	CCCGATATCC	TTACATTTCCA	GCCACATGCA	TGCGGCCCCAG	TCGAACGATA
3121	CGCAGAAGGA	GATCATCTCG	ATGGAGGACG	AAAACTCGGA	CTTGGATGAG	CCCTTCATTG
3181	ACGAGAACGG	ACCTTTTAAC	AATCTAATCT	GTTTGTTAGA	GGCCGAGAAC	GTCACTTTCC
3241	TAGCCATCTT	CCTAAACTAC	GTGATCTCAA	ACTCGGATCC	CGCGCCACTT	CTGTTTTACC
3301	TGATTACTGA	GTTGTACAAG	GAGGGCACCT	CCAAGGACAT	GCGGAAATGG	GCCTACGAAA
3361	TCCACTCCAC	ATTCCTCGTG	CCGCGGGCTC	CATTGTCTATG	GTATCGCCAA	GATGAATCGC
3421	TGGCCCCGAG	GGTGATAAT	GTCTTGCGAT	TGGAGTATGA	CAAAGTGGAG	ATCCTAAGGA
3481	CAGTTTTTCT	CCGTAGCCGA	AAGCGGGCCA	AGGACCTAAT	CAGTGAGCAG	CTGGTGTAGT
3541	TTCAGCAGAA	GCGCACCGCC	GGCCTGGGAA	CCATTTACGG	ACCCACGGAC	GACAAGCTGG
3601	CCGAGGCGAA	GACGGATAAG	CTAAGGGAGC	AAATCATCGA	CAAATATTTA	ATGCCCAATC
3661	TTCACGCGCT	TATTGAGGAT	GAGAACGGTT	CACCGCCGGA	GGATGTACGC	AAGGTGGCGT
3721	TGTGTTCTGC	TCTTTCCACC	GTCATCTACC	GTATTTTCAA	CACCTCGTCCG	CCTCCAAGCA
3781	GCATCGTTGA	GCGGGTCCAT	CACTTTGTGA	GCAGGGACAA	GAGTTTCAAG	TCGCGTATTA
3841	TGGGCAAAAA	TCGCAAGATG	AATGTTCTGT	GTCATCCATT	GGTATTGCGT	CAATACTATG
3901	AAGTGACGCA	CTGCAATCAT	TGTCAGACGA	TTATCTGGGG	CGTGAGCCCG	CAAGGTTATC
3961	ATTGTACAGA	CTGTAAATTG	AACATACACC	GTCAGTGCTC	GAAAGTAGTG	GACGAGAGTT
4021	GTCCCGGTCC	CTTGCCCCAG	GCAAAACGTC	TCGCCCACAA	CGACAAGATC	AGTAAATTCA
4081	TGGGCAAAAT	TCGACCGCGT	ACCAGCGACG	TCATTGGAAA	TGAAAAGCGA	AGTCGTCAAG
4141	ATGAGGAATT	GGATGTTGAG	TTGACTCCAG	ACCGTGGTCA	GGCGTCGATT	GTGCGTCAAC
4201	CCTCTGATCG	GCGACCGGAT	GCGAACATAT	CGATAAGGTC	GAATGGAAAT	ACCTCTTGCA
4261	ACACTTCGGG	GCTGAACACC	ACCGACCTGC	AAAGTTCTTT	TCACGGCAGC	TGTGCCAAGC
4321	ACAGTATTAA	CCCCGGCGGT	GGAGCCGGAT	GCAACATGGA	TTTATCCACG	AGCGTGGCGT
4381	CAACGACTCC	GTCGACCAGT	GGATCCGTGG	CAGCCGGTCT	GAGTGCTTTT	GCCGAACCTGA
4441	ACGCCCTGGA	TACAGTGGAT	AAAGAAGCGC	GTAGGGAGCG	TTACAGTCAG	CATCCGAAGC
4501	ACAAAAGTGC	ACCAGTCTCT	GTGAATCGGT	CGGAATCCTA	CAAGGAGCGC	TTGTCCAACA
4561	AGAGGAACCG	CAACAGTCGC	CGCAAGACCT	CTGATCCAAG	CTTGTCGTCG	CGTCCCAATG
4621	ATGAACAAC	GGACTTGGGT	CTATCGAATG	CCACTTATGT	GGGCAGTTCG	AATTCTAGTC
4681	TCTCTTCAGC	TGGCGGCAGC	GAGAGTCCCA	GCACGTCAAT	GGAGCACTTT	GCTGCACCCG
4741	GAGCAGCGGG	TGGCGTCCAG	GTGCCGCCAA	TGGGATTGAA	CCAGAACCAG	CACCCCCATC
4801	TGCTCATCCA	GCAGCACGCC	CAGCAGTACT	GCCAGCAGGA	TTCCTTTCAG	GCGGGTTTGG
4861	CAGGGGCCGC	TGGGAGCAGT	GCAGCTAGCA	ACTCTAGTTT	CTGGAATGCT	GGCCACCCAT
4921	TGCCTGTAGC	TCGTTGGACG	CTGGAGAGCG	AGGATGAAGA	CGACGTGAAC	GAGGCGGACT
4981	GGAGTTCCAT	GGTGGCCGCA	GAGGTGTTGG	CAGCCTTAAC	GGACGCTGAG	AAGAAGCGTC
5041	AGGAGATTAT	AAATGAAATC	TATCAAATCT	AACGCAACCA	TGTGCGCACC	CTAAAGCTGC
5101	TGGATCGATT	ATTCTTCCCTG	CCACTCTACG	AGAGTGGATT	GCTGTCCAG	GATCATTTGC
5161	TGTTGTTGTT	CCCCCCGCC	TTGCTGTCTG	TCCGTGAGAT	TCATGGCGCC	TTGAGCAGAA

5221 GTCTTAAGCA ACGACGCATC GAGCATAACC ACGTGGTGAA CACCATAGGG GATCTGCTCG
5281 CCGACATGTT CGATGGTCAG TCTGGAGTTG TTCTTTGCGA GTTTGCGGCT CAGTTCTGTG
5341 CCCGCCAGCA AATCGCCTTG GAGGCACTCA AGGAGAAGCG CAACAAGGAC GAGATGCTGC
5401 AAAAGTATT GAAAAAGTCG GAGTCACACA AGGCATGTCG CCGCCTCGAG CTAAAGGACT
5461 TGCTGCCCAC CGTGCTGCAG CGCCTCACCA AGTATCCGCT TTTGTTTGAG AACCTTTACA
5521 AGGTGACCGT GCGCTTGCTG CCAGAAAACA CCACCGAGGC GGAAGCCATT CAACGGGCAG
5581 TGAATCCTC TAAAAGGATT CTTGTCGAGG TCAACCAGGC AGTAAAGACA GCAGAGGATG
5641 CTCACAAGCT GCAAAACATT CAGCGTAAGT TAGACAGATC CTCTATGAC AAGGAAGAGT
5701 TTAAGAAATT GGACCTGACC CAACATCACC TCATCCATGA CGGCAATCTG ACGATCAAGA
5761 AGAATCCTAG CGTGCAGCTA CATGGACTTC TGTTTGAAAA CATGATTGTT TTGCTGACCA
5821 AGCAGGATGA TAAATATTAT CTAAAGGATT TGCACACCCC GCTATCGATC ACCAAGTAAAG
5881 CAGTCAGTCC AATTATGAGC ATTGATGCGG ACACCTTTGAT CCGGCAGGAG GCGGCTGATA
5941 AAAATTCCTT TTTCCTCATC AAGATGAAGA CATCACAAT GTTGGAGCTA CGCGCGCCTA
6001 GTAGCTCGGA GTGCAAGACA TGGTTTAAAC ACTTCTCGGA TGTGGCTGCT CGTCAGTCTA
6061 AGAATCGTTT AAAGAACGCA TCAAGCAACC ATGACACGAG TATTAGTGAT CCAGCTCTCG
6121 CCGCTATTCC GCATTCCAAC ACCAAAGAGT CGTTGGAGTT GAGCACTGAT ACAGTACAGC
6181 CATTGGCTGC GAGACCCACA TTGACTACCA CACCATTGGC CCCAATGCTG CCTATAGCCA
6241 CGTTTACACC GGTCCAGCG ACCAATAATA GTAACGTTAG CTCTCTTACT GGAGTTGAGT
6301 TGCGAAACCC TCAACGGGAT GCGACAGCAA GTGAATCTGA TGCGGATTAT GTAAACACAC
6361 CAAAGCCGCG TTCGAGCCAA AATGAAGTTA ATCGCACTAT GTCCATAAGA AGCACTGGCG
6421 AACCATTCA GAAGTATTCG GCGAATGGGA CGGAAGCAAA CGACGTTACT TTACGACACT
6481 CTCAGTCGAC TAGGGAATCG GTTAGACCAG GATCTACTGG GGAGGAGCGA AACTCCACGT
6541 ATGGTATGGT TGGAGGTAAC TCCAAACGCG ACAGCGCCAG CATTGTCTGC TCGAACAAC
6601 CGAACAACAG GCGCACCCCTT CTGATGCAGA GTCCCTTGGT GGACCCTACG GCCATTACAG
6661 TCAGCATTAG TCCCCTCAC ACAGCGGAAC CTGTGTTGAC ACCAGGAGAG AAGTTACGCC
6721 GCTTGGACGC CTCACATTAGG AATGATTTGC TGGAGAAGCA GAAAAATCATT TGTGATATCT
6781 TCCGTTTGCC AGTGGAAACAC TACGACCAGA TTGTGGACAT TGCCATGATG CCAGAGGCGC
6841 CGAAAGACAG TGCAGATATT GCTTTAGCTG CTTACGATCA GATTCAAACC CTGACCAAGA
6901 TGCTGAACGA GTACATGCAC GTCACGCCGT AGCAAGAGGT CTCAGCGGTG TCCACGGCGG
6961 TTTGTGGCCA CTGTACAGAG AAGGAAAAGC TTCGTAAGAA GGTGGCACCA TCCTCATCGT
7021 TTTCTTCATC ACCACCACCG CTGCCGCCCT CCAATAGGCA GCATGCCAG GCTCAGGCC
7081 AGATACCGCC ATCGCGGCTG ATGCCCAAAC TACAACTCT TGATCTTGAC GAAGTTGCCA
7141 TACACGAAGA CGATGACGGA TACTGTGAGA TCGACGAACT GCGCTTACCG GCTATTCCGT
7201 CCAAACCACA TGAGCGGCCC ACAACGCCAC TGGCTCCTTT CAATACTGAG CCGAAAACCT
7261 CACAATCTGT TATAGATGCC TCGAAACGTC AATCCACTGA TGCCGTTCCG GAGGGATTAC
7321 TGAACAAGA ACCACTCGAA GGCGATAAGA CGGAGACCAA GGGTGAAGAT AATGAAGTGA
7381 AAATGTGCC GTCAGATAAG CTAAGTGAAT CATGCAATGA AGAGAGGCAA TGTGTGGAGG
7441 CGGATATCAC AAAGGAAGTG GCAGATCCAA CGACCTCTAA GAATGAAGCT GCAGCATCGG
7501 TGGATGAATT ACCAAGCCAG AGCCGGGAGA TAAAAACGGC TGAAAACGCA AGCAAACTCG
7561 TAGCTGACAA AAAGGAAGAC AACGAGGAAA CCATCGAAGA AGGTGTGGCA TCCACGGTCG
7621 ATAGCTCCAC TCAAACATCA CCAACTGAAT CGCCGAAAGA GACGGATAAG TTAAGTGGAG
7681 GATCGAGCAG CACCTGTGGG CCGAATCGCA TTCAGCACGC TAGTGTGCTG GAGCCGAGTG
7741 TGCCCTGCCA TGCACTCAGC AGCATTGTAA CAATACTGAA TGAGCAGATT TCCATGCTTT
7801 TGCCAAAAAT TAACGAACGC GATATGGAAG GGGAGCGATT GCGTAAAGAG AATCAACACC
7861 TTCGCGAGCT CTTGAGTGCG CTGCATGATC GACAGCGAGT TGATGAAGTA AAGGAACTC
7921 CGTTTGATCT AAAGAAGCTG ATGCATGCTG AGGATGTAGA GTTTGACGAT GATATTGACG
7981 CCATTTCCTA CAGTTCGCTG ATGCCAACGC CTACGCCGAT TCCCACAGCA TCACCAAGCG
8041 CCAGCGGCCA GGTAGAGACA GCGGAAGCCA TGAGGATTAC TAGCACTGAG GATGAGGAAT
8101 AGGAAACTTG TTTAATTTTT TTTTGTATAA AATACATATT TGTCTTGTG TTTCCCGCAC
8161 GCATTTGACC AATTGAATTG GCGGCGTCTA TGTTTTACGA GTAGCCTTTC TTAGAGCATG
8221 CCTTTTACTA CTGTTAATGT AGCTTCATTT GTTTCTTAA AATATTATGT ATTCCACAAG
8281 AGCCGATAAC ACTTTAAATT AAGTCTAACA TTGTCCCCTC GTAAAACATA TTAATACAT
8341 ATATATATGC ATATTATCGA CTTAAAAAAT ATCAGATAAA CACACAATGT TTGATACATT
8401 TATAAAGAAC AATAAACAAAC ATAGTTATTG AGAAGAATAG CA

Protein sequences

DMCCa (630aa, 72KDa)

1	MAHPLARRPI	NVEQLSVQQL	ENRVRDLTQR	LQQAERQLTE	SNTEREICHK
50	RLEVVSQAHE	CRITEMHCVI	AELSKKLRSK	QDHVIMEEQE	PDGSELSFQE
101	GSVYNSELNL	TNPDAECQTE	PLEDFEGACS	TTSVGNVAHK	PELSHKGQV
151	EALQEEVLHL	RARIALQSE	ISTKDAAVVE	EQTKVAFDCE	SEVNECGQRL
201	NDLNVCTSLT	SPQKRIPAVP	KMAERVKLRC	ASKHESGEDP	SQDTSLSNEQ
251	INLVEHLVSE	LKEQNLVMEN	FMEPLHLSKD	LERLQRRVEQ	LEMNNTMLAL
301	TLDECKEHTE	HLYLCCGKYE	SNAVALQLAL	NCSDRAIEAY	DVMLLALLESK
351	LALLGEKSVA	AEESRRSVEA	VARHLLARLD	SEKNVCENSL	GPWQHNNINLG
401	PEDAPKTGRP	WCADDDNRLR	YHVSKLKGRR	SNVQHTIVSL	ESPFSDIYER
451	KRLALEKEHE	LRSADKKSPI	DLETAVIMQE	ILELRDSNLQ	LKTKMEEAEQ
501	ERQANANERV	ILHEALKQLQ	ANNRVSYSEA	EHAALTEQQL	VEALTRETEL
551	KGRIQTLLAN	VTASQKAFDE	KYEQLHQNV	ELQKSNHNLG	QMLDHTKRKY
601	QLRVRKLEQK	IVDLRLDYEQ	GHNHVPETTL		

DMCCb (565aa, 64KDa)

1	MHCVIAELSK	KLRSKQDHVI	MEEQEPDGSE	LSFQEGSVYN	SELNLTNPDA
51	ECQTEPLEDF	EGACSTTSVG	NVAHKPPELS	HKGQVEALQE	EVLHLRARIA
101	LLQSEISTKD	AAVVEEQTKV	AFDCESEVNE	CGQRLNDLNV	CTSLTSPQKR
151	IPAVPKMAER	VKLRCASKHE	SGEDPSQDTS	LSNEQINLVE	HLVSELKEQN
201	LYMENFMEPL	HLSKDLERLQ	RRVEQLEMNR	TMLALTLEDC	KEHTEHLYLL
251	CGKYESNAVA	LQALNCSDR	AIEAYDVMLA	LLESKLALLG	EKSVAAEESR
301	RSVEAVARHL	LARLDSEKNV	CENSLGPWQH	NINLGPEDAP	KTGRPWCADD
351	DNRLRYHVS	LKGRRSNVQH	TIVSLESPPS	DIYERKRLAL	EKEHELRSAD
401	KKSPIDLETA	VIMQEILELR	DSNLQLKTKM	EEAEQERQNA	NERVGILHEA
451	LKQLQANNRV	SYSEAEHAAL	TEQQLVEALT	RETELKGRIQ	TLLANVTASQ
501	KAFDEKYEQL	HQNVRELQKS	NHNLGQMLDH	TKRKYQLRVR	KLEQKIVDLR
551	LDYEQGHNV	PETTL			

DMCCc (718aa, 82KDa)

1	MSNDVQVQAR	VAKIATDVPR	RSQKQRDSSQ	FQQKHSQSAA	GEDFEYVFQS
51	ISPRGGGPGP	GGRHLVGSSD	LDSPEHTQRD	TTESDNNISS	CSTLDIVNKV
101	EQLSVQQLEN	RVRDLTQRLQ	QAERQLTESN	TEREICHKRL	EVVSQAHECR
151	ITEMHCVIAE	LSKKLRSKQD	HVIMEEQEPD	GSELSFQEGS	VYNSELNLTN
201	PDAECQTEPL	EDFEGACSTT	SVGNVAHKPP	ELSHKGQVEA	LQEEVLHLRA
251	RIALLQSEIS	TKDAAVVEEQ	TKVAFDCESE	VNECGQRLND	LVNCTSLTSP
301	QKRIPAVPKM	AERVKLRCAS	KHESGEDPSQ	DTLSLSNEQIN	LVEHLVSELK
351	EQNLVMENFM	EPLHLSKDLE	RLQRRVEQLE	MRNTMLALTL	DECKEHEHL
401	YLLCGKYESN	AVALLQALNC	SDRAIEAYDV	MLALLESKLA	LLGEKSVAEE
451	ESRRSVEAVA	RHLLARLDSE	KNVCENSLGP	WQHNNINLGPE	DAPKTGRPWC
501	ADDDNRLRYH	VSKLKGRRSN	VQHTIVSLES	PFSDIYERKR	LALKEHEHLR
551	SADKKSPIDL	ETAVIMQEIL	ELRDSNLQLK	TKMEEAEQER	QNANERVGIL
601	HEALKQLQAN	NRVSYSEAEH	AALTEQQLVE	ALTRETELKG	RIQTLLANVT
651	ASQKAFDEKY	EQLHQNVREL	QKSNHNLGQM	LDHTKRKYQL	RVRKLEQKIV
701	DLRLDYEQGH	NHVPETTL			

DMCCd (337aa, 37KDa)

1	MSNDVQVQAR	VAKIATDVPR	RSQKQRDSSQ	FQQKHSQSAA	GEDFEYVFQS
51	ISPRGGGPGP	GGRHLVGSSD	LDSPEHTQRD	TTESDNNISS	CSTLDIVNKV
101	EQLSVQQLEN	RVRDLTQRLQ	QAERQLTESN	TEREICHKRL	EVVSQAHECR
151	ITEMHCVIAE	LSKKLRSKQD	HVIMEEQEPD	GSELSFQEGS	VYNSELNLTN
201	PDAECQTEPL	EDFEGACSTT	SVGNVAHKPP	ELSHKGQVEA	LQEEVLHLRA

251 RIALQLQSEIS TKDAAVVEEQ TKVAFDCESE VNECGQRLND LNVCTSLTSP
 301 QKRIPAVPKM AERVKLRCAS KHESGEDPSQ DTSLNSNE

HsMCC1 (829aa, 93 KDa)

1 MNSGVAMKYG NDSSAELSEL HSAALASLKG DIVELNKRLQ QTERERDLLE KKLAKAQCEQ
 61 SHLMREHEDV QERTTLRYEE RITELHSVIA ELNKKIDRLQ GTTIREEDEY SELRSELSQS
 121 QHEVNEDSRS MDQDQTSVSI PENQSTMVTA DMDNCSLNS ELQRVLTGLE NVVCGRKKSS
 181 CSLSVAEVDR HIEQLTTASE HCDLAIKTVE EIEGVLGRDL YPNLAEERSR WEKELAGLRE
 241 ENESLTAMLC SKEEELNRTK ATMNAIREER DRLRRRVREL QTRLQSVQAT GPSSPGRLTS
 301 TNRPINPSTG ELSTSSSSND IPIAKIAERV KLSKTRSESS SSDRPVLGSE ISSIGVSSSV
 361 AEHLAHSQD CSNIQEIFQT LYSHGSAISE SKIREFEVET ERLNSRIEHL KSQNDLLTIT
 421 LEECKSNAER MSMLVGKYES NATALRLALQ YSEQCIEAYE LLLALAESEQ SLILGQFRAA
 481 GVGSSPGDQS GDENITQMLK RAHDCRKTAE NAAKALLMKL DGSCGGAFAV AGCSVQPWES
 541 LSSNSHTSTT SSTASSCDTE FTKEDQRLK DYIQQLKNDR AAVKLTMLEL ESIHIDPLSY
 601 DVKPRGDSQR LDLENVLMQ ELMAMKEEMA ELKAQLYLLE KEKKALELKL TREAEQAY
 661 LVHIEHLKSE VEEQKEQRM RLSSTSSGSK DKPGKECADA ASPALSLAEL RTTCSNELA
 721 AEFTNAIRRE KKLKARVQEL VSALERLTKS SEIRHQQSAE FVNDLKRANS NLVAAEYKAK
 781 KKHQNKLLKL ESQMMAMVER HETQVRMLKQ RIALLEEENS RPHTNETSL

HsMCC2, AIE-75 binding protein (723aa, 76Kda)

1 MSARATRP RS RGRHAPPGE LDPVAESSEE VEAASGSSKP SFAPPPVSSG LEQLGPMEEV
 61 SGQGLGSR TD KKMDGGSGRE LASAPEVPHK PAVEAHQAPE AALQYKETVP PGNGAPDV FQ
 121 TLQHTLSS LE AAAAARHQP PSHSGPMEFE GTSEGGAGSL GKQEGAGSCQ REAARLAERN
 181 AWLRLALSS R EDELVRTQAS LEAIRAEKET LQKEVQELQD SLLRLEPCPH LSHNQAGGSG
 241 SGSSSEADR EPWETQDSFS LAHPLLRLR SHSSTQILGS LPNQPLSP EM HIMEAQMEQL
 301 RGSIEKLKCF NRLLSAVLQG YKGRCEGLSM QLGQREAEAT ALHLALQYSE HCEEAYRVLL
 361 ALREADSGAG DEAPMSDLQA AEKEAWRLLA QEEAAMDAGA QQNPPQSP EG SSVDKPTPQE
 421 VAFQLRSYVQ RLQERRSLMK ILSEPGPTLA PMPTVPRAEA MVQAILGTQA GPALPRLEKT
 481 QIQQDLVAAR EALADLMLRL QLVRREKRGL ELREAALRAL GPAHVLLLEQ LRWERAELQA
 541 GGANSSGGHS SGGGSSGDEE EWYQGLPAVP GGTSGIDGGQ VGRAWDPEKL AQELAASLTR
 601 TLDLQEQ LQS LRRELEQVAQ KGRARRSQA ELNRDLCKAH SALVLAFRGA HRKQEEQRRK
 661 LEQQMALMEA QQAEEVAVLE ATARALGKPR PPLPPPQLGD
 721 TFL

DRho GEF2 (2559aa, 281KDa)

1 MDDPSIKKRL LDLYTDEHEY DEVQEIP EES SIQPPETSTS HTSTNGSSHS GPGTATGPGA
 60 TSAGPSAGAP QSPVIVVDSV PELPAPKQKS VKNSKSKQKQ KQLANKSKIP RSPSLASSLS
 120 SLASSLSGHR DRDKDRDKDR ENQNAVPPQT PPLPPSYKQN QMNGDSTAAA GGGVSAPATP
 180 TTANNNNASH NNGSIMGGGV QLNQSDNSNP VLQAPGERSS LNLTPLSRDL SGGHTQESTT
 240 PATTPSTPSL ALPKNFQYLT LTVRKDSNGY GMKVSGDNPV FVESVKPGGA AEIAGLVAGD
 300 MILRVNGHEV RLEKHPTVVG LIKASTVEL AVKRSQKLTR PSSVSVVTPS TPILSGDRDT
 360 ASITGPQ PVD SIKRREMETY KIQT LQKMLE QEKLNLERLK SDQNNPSYKL SEANIRKLRE
 420 QLHQVGAEDA PTVKLQAAAG NKNTALLTPN QIQHLSASAT HSNQQFHH LH HHHNLHNNNY
 480 PPQQQPASTS PAFLSLLPRS LSSLSLGRK NKTEKDLTTS SPFGLTDFL QQQRMSHQAE
 540 SMSQSMHQHT STPTSQQFFH PHQQQHRFKE TGPTSKGKNK FLISRSLIEE DVPPPLPQRN
 600 PPRQLNLDLK NGNASPGGSH LVAPVSD LDR ATSPQLNRSQ QQQLPRSTDN SPSNAKSKRS
 660 KIKTKALSDP KMSTQMFLQM ESASAAGAAG GSIEVDGGPP PLPPRLPGMM TEDMSRGSCQ
 720 NLAQPNSVGT AFNYPLVSTT TAVQNDNLNI AFPLSQRPN I VQQLQQYQQQ QQHQMSGGQA
 780 TGALQQT PNL GKNKHRRVGS SPDNMHPRHP DRITKTTS GS WEIVEKDGES SPPGTPPPPY
 840 LSSSHMTVLE PDNENNRGAA AAGPGVFIES HQFTPMAGAS SPIPISLHSN HMHAAQSDNT
 900 QKEIISMEDE NSDLDEPFID ENGPFFNLTR LLEAENVTF L AIPLNYVISN SDPAPLLFYL
 960 ITELYKEGTS KDMRKWAYEI HSTFLVPRAP LSWYRQDES L AREVDNLVQL EYDKVEILRT
 1020 VFLRSRKR AK DLISEQLREF QQKRTAGLGT IYGPTDDKLA EAKTDKLREQ IIDKYLMPNL
 1080 HALIEDENG S PPEDVRKVAL CSALSTVIYR IFNTRPPSS IVERVHHFVS RDKSFKSRIM

1140	GKNRKMNVRG	HPLVLRQYYE	VTHCNHCQTI	IWGVSPQGYH	CTDCKLNIHR	QCSKVVDSC
1200	PGPLPQAKRL	AHNDKISKFM	GKIRPRTSDV	IGNEKRSRQD	EELNVELTPD	RGQASIVRQP
1260	SDRRPDANIS	IRSNGNTSCN	TSGLNTTDLQ	SSFHGSCAND	SINPGGGAGC	NMDLSTSVAS
1320	TTPSTSGSVA	AGLSAFAELN	ALDTVDKEAR	RERYSQHPKH	KSAPVSVNWS	ESYKERLSNK
1380	RNRNSRRKTS	DPSLSSRPND	EQLDLGLSNA	TYVGSSNSSL	SSAGGTESPS	TSMEHFAAPG
1440	AAGGVQVPPM	GLNQNQHPHL	LIQQHAQQYC	QQDSFQAGLA	GAAGSSAASN	SSFWNAGHPL
1500	PVARWTLESE	DEDDVNEADW	SSMVAAEVSA	ALTDAEKKRQ	EIINEIYQTE	RNHVRTLKL
1560	DRLFFLPLYE	SGLLSQDHLL	LLFPPALLSL	REIHGAFEQS	LKQRRIEHNH	VVNTIGDLLA
1620	DMFDGQSGVV	LCEFAAQFCA	RQQIALEALK	EKRNKDEMLQ	KLLKKSESHK	ACRRLELKDL
1680	LPTVLQRLTK	YPLLFENLYK	VTVRLLPENT	TEAEAIQRAV	ESSKRILVEV	NQAVRTAEDA
1740	HKLQNIQRKL	DRSSYDKEEF	KKLDLTQHHL	IHDGNLTIKK	NPSVQLHGLL	FENMIVLLTK
1800	QDDKYLYKNL	HTPLSITNKP	VSPIMSIDAD	TLIRQEADK	NSFFLIKMK	SQMLELRAPS
1860	SSECKTWFKH	FSDVAARQSK	NRSKNASSNH	DTSISDPALA	AIPHSNTKES	LELSTDTVQP
1920	LAATATLTTT	PLAPMLPIAT	VTPAPATNNS	NVSSLTGVQL	RNPQRDATAS	ESDADYVNTF
1980	KPRSSQNEVN	RTMSIRSTGE	PIQKYSANGT	EANDVTLRHS	QSTRESVRPG	STGEERNSTY
2040	GMVGGNSKRD	SASIVCSNNS	NNTRTLLMQS	PLVDPTAIQV	SISPAHTAEP	VLTPGEKLRR
2100	LDASIRNDLL	EKQKIICDIF	RLPVEHYDQI	VDIAMMPEAP	KDSADIALAA	YDQIQTLTKM
2160	LNEYMHVTPE	QEVSAVSTAV	CGHCHEKEKL	RKKVAPSSSF	SSSPPLPPP	NRQHAQAQAQ
2220	IPPSRLMPKL	QTLDLDEVAI	HEDDDGYCEI	DELRLPAIPS	KPHERPTTPL	APFNTEPKTS
2280	QSVIDASKRQ	STDAVPEGLL	EQEPLEGDKT	ETKGEDNEVK	TVPSDKLSES	CNEERQCVEA
2340	DITKEVADPT	TSKNEAAASV	DELPSQSREI	KTAENASKSV	ADKKEDNEET	IEEGVASTVD
2400	SSTQTSPTES	PKETDKLTGG	SSSTCGPNRI	QHASVLEPSV	PCHALSSIVT	ILNEQISMLL
2460	PKINERDMER	ERLRKENQHL	RELLSALHDR	QRVDEVKETP	FDLKKLMHAE	DVEFDDDDIDA
2520	ISNSSLTPTP	TPIPTASPSA	SGQVETAEAM	RITSTEDEE		

Appendix 2: Primer sequences

In situ_1	TCTAGATGAGAGCAACACAGAGCG
In situ_2	CTCGAGGCAATTAAGTCGCATGGC
MCC RT PCR_1	GCCAATTCTAGTCGGAAAGTCGTTGCA
MCC RT PCR_2	GATACGCCATCGGTGCCCAACGACAA
MCC RT PCR_3	GGGTTGTATGGGTTGAAATGTATTCTCTGA
MCC RT PCR_4	GTATGTAGTTACTGCAGCCTTATCTGCG
MCC RT PCR_5	ACTTGATCATGTTTAGTACCAATGTATCATTGC
MCC RT PCR_6	CGAGGGGCTTCTACTAATGGAAGTCCA
MCC RT PCR_7	GGTGTCTGGAGCTAAAGGAGCAAAATCT
cDNA_for	GCTCTAGAAATCGGAAGTCAACGAGT
cDNA_rev	CAGAAATTCATTGCTTTCGAGCAGAGC
cDNA_int_for	GCACTGCAGCTAGACGAATGCAAG
cDNA_int_rev	TAGCTGCAGTGCCAGCATGGTGTT
Gen_for	TGCGGGCCGCAATCGGAAGTCAACGAGT
Gen_rev	CCCGAATTCTGCCAACCTGGAAGTGATG
PlacW5'	TCCTCTCAACAAGCAAACGTGCACTG
PlacW3'	CGCTGTCTCACTCAGACTCAATACGA
G8	TACAGCCGATAGTGGCGCTAAACGAA
G9	TTCGTTTAGCGCCACTATCGGCTGTA
G10	CCCAAATCAGGCCAATTCTAGTCGGA
G11	TCCGACTAGAATTGGCCTGATTTGGG
Pp3	ATCCGGCCGCTGTCATTTCTTGTTT
Taq1	TAGATTTTGCTCCTTTAGCTCCGACAC
Taq2	GTGTCGGAGCTAAAGGAGCAAAATCTA
Taq3	CCACGCCCATAAAGAACTACTCGATG
Taq4	TAAGCCCGTCTAGGTTGTGATAATGTTGT
3'gen	ATCATGCATTGCCTTAGGGCTGGAATTTA
5'gen	TAATAGCGAACTTCTGGCTGTCGGT
LacW5'probe1	TTCATCCACCACATACAGGCCGTAG
LacW'5'probe2	CAGTGCACGTTTGCTTGTTGAGAGGA
LacW3'probe1	TCGTATTGAGTCTGAGTGAGACAGCG
LacW3'probe2	AAGATGCTGAAGATCAGTTGGGTGCAC
Pel2.8out	CGGTTAAATTGCCAACGCTTATTACCC
Pel3.8out	TGCTGATATGGTTGATGTCATGTAGCC
Pel4.8out	CTCCGCGAATTAATAGCTCCTGATC
Pel5.8out	AAAGAGGTATCCTGCTGGACATAG
Pel6.8out	GGCAACCATCTGCAAATTTAAATGTTACTCG
Pel7.8out	GTTCGAAGGGGAAATACTTGTATTCTATAGG
Pel8.8out	GTGCACCCAACCTGATCTTCAGCATCTT
Pel9.5out	CTGACAGTTACCAATGCTTAATCATGTAG
Pel10.2out	TTGCGTTGCTGGCGTTTTTCCATAGGC
Pel1.8in	CTACGGCCTGTATGTGGTGGATGAA
Pel3.8in	GGCTACATGACATCAACCATATCAGCA
Pel7.8in	CCTATAGAATACAAGTATTTCCCCTTCGAAC
Gen4Kb	CACATAGCTCGAACATCGTGTGGGAA
Gen8Kb	CCGAGGTGTAGTTGGAGCCATTCATTA
Gen12Kb	TATTTGCCGCCCATCCAATTCACACA
Gen16Kb	TATGGCACAATACTGACTGAGGTCCT

Gen20Kb	ATGTCTCTCATTTGGGGTGGGCTC
Gen24Kb	TATTGCGTTGTAACCACTGACACTGAC
Gen28Kb	TTTCAACTTCCAGTCGTATCCCCGC
Gen4Kb2	TTCCCACACGATGTTTCGAGCTATGTG
Gen8Kb2	TAATGAATGGCTCCAACTACACCTCGG
Gen12Kb2	TGTGTGAATTGGATGGGCGGCAAATA
Gen16Kb2	AGGACCTCAGTCAGTATTGTGCCATA
Gen20Kb2	GAGCCCACCCCAAATGAGAGACAT
MyctagA5'	GATCCAACATGGAGCAGAAGCTGATCAGCGAAG
(cont'd)	AGGATCTG
MyctagA3'	AATTCAGATCCTCTTCGCTGATCAGCTTCTGCT
(cont'd)	CCATGTTG
MyctagB5'	GATCCAACATGGAGCAGAAGCTGATCAGCGAAG
(cont'd)	AGGATCTGG
MyctagB3'	AATTCCAGATCCTCTTCGCTGATCAGCTTCTGC
(cont'd)	TCCATGTTG
MyctagC5'	GATCCAACATGGAGCAGAAGCTGATCAGCGAAG
(cont'd)	AGGATCTGAG
MyctagC3'	AATTCTCAGATCCTCTTCGCTGATCAGCTTCTG
(cont'd)	CTCCATGTTG

Appendix 3: TXL-COOH genes of *Drosophila* (PATSCAN result)

<u>Gene</u>	<u>C-terminus</u>	<u>Gene</u>	<u>C-terminus</u>	<u>Gene</u>	<u>C-terminus</u>
CG10002	TSL	CG2189	TAL	CG7248	TEL
CG10194	TKL	CG2595	THL	CG7285	TDL
CG10210	TEL	CG2849	TLL	CG7635	TNL
CG10405	TKL	CG30033	TAL	CG7919	TRL
CG10413	TTL	CG30361	TNL	CG7997	TAL
CG10420	TEL	CG30479	TIL	CG8085	TRL
CG10493	TEL	CG31009	TEL	CG8261	TVL
CG10874	TGL	CG31043	TAL	CG8484	TYL
CG10933	TSL	CG31066	TRL	CG8507	TEL
CG11136	TEL	CG31100	THL	CG8759	TML
CG11282	TEL	CG31321	TSL	CG8856	TDL
CG11560	TNL	CG3204	TLL	CG8983	TEL
CG11598	TFL	CG32045	TNL	CG9169	TFL
CG11898	TTL	CG32089	TKL	CG9240	TGL
CG12724	TNL	CG32096	TDL	CG9302	TEL
CG12928	TTL	CG32220	TEL	CG9512	TEL
CG12931	TIL	CG32387	TYL	CG9610	TGL
CG12991	TKL	CG32578	TTL	CG9764	TQL
CG13068	TVL	CG32594	TDL	CG9864	TKL
CG13566-	TDL	CG32711	TFL	CG9993	THL
CG13702	TDL	CG3318	TEL	CG9997	TIL
CG13711	TEL	CG3389	TEL	CG1907	TGL
CG13949	TDL	CG3638	TEL	CG7218	TQL
CG14112	TTL	CG3777	TIL		
CG14117	TKL	CG3897	TEL		
CG14427	TDL	CG4099	TDL		
CG14669	TIL	CG4301	TYL		
CG1467	TKL	CG4376	TDL		
CG14723	TFL	CG4462	TRL		
CG14973	TNL	CG4626	TLL		
CG15097	TNL	CG4889	TCL		
CG15134	THL	CG4928	TRL		
CG15153	TDL	CG4969	TCL		
CG15214	TLL	CG5235	TLL		
CG15377	TFL	CG5270	TPL		
CG1636	TAL	CG5507	TEL		
CG16752	TVL	CG5539	TGL		
CG1693-	TEL	CG5671	TYL		
CG17084	TSL	CG5874	TIL		
CG17131	TGL	CG6156	TTL		
CG1722	THL	CG6477	TAL		
CG17389	TIL	CG6600	TKL		
CG18404	TNL	CG6619	TRL		
CG18572	TAL	CG6805	TQL		
CG18745	TWL	CG6988	TPL		
CG18746	TWL	CG7128	TNL		